

UNIVERSITÉ DE SHERBROOKE
Faculté de génie
Département de génie électrique et de génie informatique

Récupération en temps réel de coïncidences
diffuses triples dans un scanner TEP à l'aide
d'un réseau de neurones artificiels

Mémoire de maîtrise
Spécialité : génie électrique

Charles Geoffroy

Jury : Réjean Fontaine (directeur)
Charles-Antoine Brunet (co-directeur)
Frédéric Mailhot
Daniel Dalle

Sherbrooke (Québec) Canada

Juillet 2013



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-499-00328-7

Our file Notre référence

ISBN: 978-0-499-00328-7

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

RÉSUMÉ

Le projet de recherche s'inscrit dans un contexte d'imagerie moléculaire, où la modalité d'imagerie d'intérêt est la tomographie d'émission par positrons (TEP) appliquée en recherche sur les petits animaux. Afin de permettre l'observation de détails infimes, les plus récents développements sur ce genre de scanner ont constamment amélioré leur résolution spatiale, sans toutefois obtenir les mêmes progrès en terme de sensibilité. Parmi les méthodes étudiées afin de combler cette lacune, la récupération de coïncidences triples à l'aide d'un réseau de neurones artificiels semble être une technique viable. En effet, malgré une dégradation du contraste, celle-ci permet d'améliorer substantiellement la sensibilité de l'image. Cette technique n'est cependant pas prête à être intégrée aux protocoles de recherche, car son application est pour l'instant limitée à un traitement hors ligne des données d'acquisition d'un scanner. En conséquence, la faisabilité d'une telle approche en temps réel n'est donc pas garantie, car le flux de coïncidences d'un scanner est très important et ses ressources de calculs sont limitées.

Dans l'intention d'inclure ce gain en sensibilité pendant une acquisition où le traitement est effectué en temps réel, ce projet de recherche propose une implémentation d'un réseau de neurones artificiels au sein d'une matrice de porte programmable (FPGA) pouvant récupérer en temps réel les coïncidences diffuses triples du scanner LabPET, version 4 cm. La capacité de traitement obtenue est 1 087 000 coïncidences triples par seconde en utilisant 23.1% des ressources d'unités logiques d'un FPGA de modèle XC2VP50. Comparativement à un programme équivalent à haute précision sur ordinateur personnel, l'analyse de validité prend la même décision dans 99.9% des cas et la même ligne de réponse est choisie dans 97.9% des cas. Intégrées à l'image, les coïncidences triples permettent une augmentation de sensibilité jusqu'à 39.7%, valeur qui est en deçà celle obtenue des recherches antérieures, mais expliquée par des conditions d'acquisition différente. Au niveau de la qualité de l'image, la dégradation du contraste de 16,1% obtenu est similaire à celle observée antérieurement. En référence à ces résultats, les ressources limitées d'un scanner de tomographie d'émission par positrons sont avérées suffisantes pour permettre l'implémentation d'un réseau de neurones artificiels devant classer en temps réel les coïncidences triples du scanner.

En terme de contributions, l'implémentation en temps réel réalisée pour ce projet confirme la faisabilité de la technique et apporte une nouvelle approche concrète pour améliorer la sensibilité. Dans une autre mesure, la réussite du projet de recherche contribue à faire connaître la technique des réseaux de neurones artificiels dans le domaine de la tomographie d'émission par positrons. En effet, cette approche est pertinente à considérer en guise d'alternative aux solutions traditionnelles. Par exemple, les réseaux de neurones artificiels pourraient effectuer une évaluation correcte du phénomène des coïncidences fortuites.

Mots-clés : Tomographie d'émission par positrons, réseau de neurones artificiels, FPGA, coïncidences diffuses triples, sensibilité en TEP

TABLE DES MATIÈRES

1	Introduction	1
1.1	Mise en contexte et problématique	1
1.2	Définition et objectifs du projet de recherche	3
1.3	Contributions originales	4
1.4	Plan du document	4
2	Tomographie d'émission par positrons	7
2.1	Description générale	7
2.1.1	Principes de base	7
2.1.2	Interactions des photons d'annihilation avec la matière	8
2.1.3	Types de coïncidences	9
2.2	Défis de la TEP destinée aux petits animaux	10
2.3	Résolution spatiale	10
2.3.1	Facteurs physiques	11
2.3.2	Dimension des détecteurs	11
2.3.3	Localisation des détecteurs	12
2.4	Sensibilité	13
2.4.1	Facteurs de sensibilité	13
2.4.2	Temps mort	13
2.5	Contraste	14
2.5.1	Résolution en temps	14
2.5.2	Résolution en énergie	15
2.6	Tendance des dernières années	15
3	Amélioration de la sensibilité d'un scanner	17
3.1	Équations de Klein-Nishina	17
3.2	Classification bayésienne	18
3.3	Réseaux de neurones artificiels	19
4	Réalisation d'un RNA dans un FPGA	23
4.1	Faisabilité	23
4.2	Représentation des données	23
4.3	Représentation des fonctions complexes	24
4.4	Implémentation d'un réseau de neurones artificiels	25
5	Intégration au scanner LabPET™	27
5.1	Architecture du LabPET™	27
5.2	Acquisition des événements	28
5.3	Formation des coïncidences	29
5.4	Collimation électronique	30
5.5	Extraction des données	31

5.6	Insertion de la récupération des coïncidences triples	32
5.7	Respect des contraintes	34
5.7.1	Contraintes de performance	34
5.7.2	Contraintes de ressources	34
6	Conception de la récupération des coïncidences triples	37
6.1	Architecture	37
6.2	Représentation des données	38
6.3	Contrôle des entrées	39
6.4	Validation d'une coïncidence triple	41
6.5	Simplification des données	42
6.5.1	Calcul de l'angle de rotation par arc-tangente	42
6.5.2	Rotation par quaternion	43
6.5.3	Calcul des composantes de rotation par cosinus et sinus	45
6.5.4	Résultats	45
6.6	Normalisation des données	51
6.6.1	Multiplication inverse	51
6.6.2	Calcul de la normalisation	52
6.6.3	Résultats	52
6.7	Réseau de neurones artificiels	54
6.7.1	Blocs de calcul	55
6.7.2	Poids et biais	56
6.7.3	Fonction d'activation tangente hyperbolique	56
6.7.4	Assignation des entrées des neurones	57
6.7.5	Bloc de contrôle	57
6.7.6	Résultats	57
6.8	Contrôle des sorties	59
7	Méthodologie	61
7.1	Acquisition des données	61
7.2	Filtrage des données	62
7.3	Reconstruction d'image	62
7.3.1	Isocompte	63
7.3.2	Isolongueur	63
7.4	Évaluation sur l'image	64
7.4.1	Sensibilité	64
7.4.2	Contraste	64
7.4.3	Ligne de profil	65
8	Résultats	67
8.1	Acquisition des données	67
8.2	Filtrage des données	67
8.3	Reconstruction d'image	69
8.3.1	Isocompte	69
8.3.2	Isolongueur	71

TABLE DES MATIÈRES	v
8.4 Évaluation sur l'image	73
8.4.1 Sensibilité	73
8.4.2 Contraste	74
8.4.3 Ligne de profil	75
8.5 Discussion	77
8.5.1 Évaluation des résultats	77
8.5.2 Acquisitions à activité radioactive plus intense	78
8.5.3 Optimisations de performance et de ressource	79
8.5.4 Compatibilité aux autres versions du LabPET	80
9 Conclusion	83
9.1 Sommaire	83
9.2 Contributions	84
9.3 Travaux futurs	84
A Simplification des quaternions	87
B Valeurs de poids et biais	91
C Valeur du champ Baseline	93
LISTE DES RÉFÉRENCES	95

LISTE DES FIGURES

2.1	Émission de positron	7
2.2	Coïncidences en TEP	9
2.3	Effet parallaxe	12
3.1	Réseau de neurones artificiels	20
5.1	Géométrie du LabPET TM	27
5.2	Groupement des événements en coïncidences	29
5.3	Chaînes de formation des coïncidences	29
5.4	Collimation électronique radiale d'une coïncidence double	30
5.5	Collimation électronique radiale d'une coïncidence double	30
5.6	Collimation électronique radiale d'une coïncidence triple	31
5.7	Insertion du traitement avant la collimation électronique	32
5.8	Détour complet des coïncidences triples	33
5.9	Insertion de la récupération des coïncidences triples	33
6.1	Architecture de la récupération des triples	37
6.2	Registre à décalage de la propagation des signaux de départ	40
6.3	Calcul des équations 6.12 et 6.13 du quaternion	44
6.4	Suppression des symétries : histogramme des différences en X	46
6.5	Suppression des symétries : histogramme des différences en Y	47
6.6	Alignement du point milieu à $Y = 0$: histogramme des différences en X	47
6.7	Alignement du point milieu à $Y = 0$: histogramme des différences en Y	48
6.8	Alignement du point milieu à $Z = 0$: histogramme des différences en X	48
6.9	Alignement du point milieu à $Z = 0$: histogramme des différences en Z	49
6.10	Suppression composante Z : histogramme des différences en Y	49
6.11	Suppression composante Z : histogramme des différences en Z	50
6.12	Normalisation : histogramme des différences en X	53
6.13	Normalisation : histogramme des différences en Y	53
6.14	Normalisation : histogramme des différences en énergie	54
6.15	Architecture du réseau de neurones artificiels	55
6.16	RNA : histogramme des différences	58
7.1	Régions utilisées pour l'évaluation du contraste	64
7.2	Sélection du CNR en fonction du nombre d'itérations	65
7.3	Ligne de profil utilisée	65
8.1	Lignes de profil des images de type isocompte	75
8.2	Lignes de profil des images de type isolongueur	76

LISTE DES TABLEAUX

2.1	Performances des scanners TEP pour petits animaux	15
5.1	Représentation du paquet d'un événement	28
5.2	Ressources disponibles de l'engin de coïncidence du LabPET4	35
6.1	Ressources utilisés dans l'engin de coïncidence LabPET4	37
6.2	Évaluation des étendues des étapes préalables au RNA	38
6.3	Évaluation des étendues du RNA	39
6.4	Caractéristiques des représentations de données choisies	39
6.5	Ressources utilisées par le contrôle des entrées	40
6.6	Ressources utilisées par le module de validation	42
6.7	Assignation des entrées pour une rotation par quaternion.	44
6.8	Assignation des sorties pour une rotation par quaternion.	44
6.9	Assignation des entrées et sorties	46
6.10	Ressources utilisées par les fonctions principales	50
6.11	Ressources utilisées les modules de simplification des données	51
6.12	Assignation des entrées et sorties	52
6.13	Ressources utilisées par la normalisation des données	54
6.14	Multiplexeurs contrôlant les entrées des neurones	57
6.15	Signaux de contrôle du RNA	58
6.16	Ressources utilisées par le réseau de neurones artificiels	59
6.17	Reconfiguration du champ baseline	59
6.18	Ressources utilisées par le contrôle des sorties	60
7.1	Révisions utilisés pour l'acquisition	61
7.2	Coïncidences incluses dans chacune des images isolongueur	63
8.1	Total des coïncidences obtenues	67
8.2	Coïncidences après l'application d'un seuil de 125 keV	67
8.3	Coïncidences pouvant être intégrées à l'image	68
8.4	Total des coïncidences obtenues	73
8.5	Rapports de contraste sur bruit obtenus (Isocompte)	74
8.6	Rapports de contraste sur bruit obtenus (Isolongueur)	74
B.1	Valeurs de poids et biais.	92
C.1	Valeurs du champ Baseline	93

LISTE DES ACRONYMES

Acronyme	Définition
CNR	Contrast-to-noise ratio
CT	Computed Tomography
CZT	Cadmium Zinc Telluride
D250	Doubles valides au seuil d'énergie de 250 keV
D360	Doubles valides au seuil d'énergie de 360 keV
FPGA	Field-Programmable Gate Array
FWHM	Full Width Half Max
IEEE	Institute of Electrical and Electronics Engineers
IRM	Imagerie par résonance magnétique
IRMd	Imagerie par résonance magnétique fonctionnelle
LGSO	$Lu_{0,4}Gd_{1,6}SiO_5$
LUT	Look-up table
LYSO	$Lu_{1,8}Y_{0,2}SiO_5(Ce)$
MAP	Maximum a posteriori
MIC	Medical Imaging Conference
NSS	Nuclear Science Symposium
PDA	Photodiode avalanche
RMS	Root Mean Square
RNA	Réseau de neurones artificiels
SPECT	Single-Photon Emission Computed Tomography
THL	Triples récupérés par une technique basée sur l'énergie
TEP	Tomographie d'émission par positrons
TI	Triples considérés invalides
THL	Triples récupérés par traitement hors-ligne
TTR	Triples récupérés en temps réel
VHDL	VHSIC Hardware Description Language

CHAPITRE 1

Introduction

1.1 Mise en contexte et problématique

Dans sa pratique actuelle, la plupart des diagnostics et traitements de la médecine moderne ne sont effectués qu'à partir du début des symptômes d'une maladie, ce qui est parfois bien après le début des premiers changements anatomiques non visuellement perceptibles. Suite à ce délai, la progression d'une maladie peut souvent atteindre un point où les traitements appliqués sont plus difficiles, plus coûteux et où les chances de guérison sont amoindries. C'est pourquoi dans les dernières décennies, un effort particulier a été investi dans l'imagerie moléculaire. Effectivement, les avancées technologiques réalisées dans ce domaine ouvrent aujourd'hui la voie à une panoplie de nouvelles procédures diagnostiques et thérapeutiques ciblées pouvant avoir un impact direct sur les soins apportés à un patient.

De façon générale, l'imagerie moléculaire réfère à la mesure et à la caractérisation *in vivo* des processus biologiques situés au niveau cellulaire et moléculaire [65]. Contrairement aux méthodes plus classiques d'imagerie diagnostique, cette dernière se démarque avec sa capacité d'observer les anomalies moléculaires à l'origine d'une maladie plutôt que d'en constater les effets anatomiques. Par cette habilité, l'imagerie moléculaire permet non seulement une détection plus rapide d'une maladie, mais a aussi le potentiel d'imager les progrès d'un traitement. Par ailleurs, avec la récente découverte de la séquence du génome humain, cette technique permet à la recherche d'identifier de nouvelles cibles pour une détection plus rapide des maladies, d'imager l'expression des gènes, d'évaluer de nouveaux traceurs moléculaires, etc.

L'imagerie moléculaire est constituée de plusieurs modalités d'imagerie, où chaque technique différente emploie une stratégie unique afin de produire un signal servant à la reconstruction d'une image. Parmi celles-ci, il est possible de produire une image du magnétisme des variations hémodynamiques (IRMf), de la lumière (bioluminescence ou fluorescence), ou même par des désintégrations radioactives (SPECT, TEP). Cette variété de stratégies possibles est très importante, car elle permet de tirer profit des avantages de chacune des techniques de façon à améliorer la recherche sur les processus moléculaires.

Parmi les techniques d'imagerie moléculaire, la TEP, dérivée de la médecine nucléaire, permet d'imager la distribution radioactive de la zone étudiée [30]. Cette capacité quantitative lui procure alors la possibilité de produire des évaluations fonctionnelles, comparativement à d'autres modalités d'imagerie plus spécialisées à fournir un portrait anatomique. En d'autres mots, la TEP révèle ce que les cellules font au sein des tissus ou des organes. D'un point de vue plus applicatif, la TEP est particulièrement présente en oncologie et en cardiologie [81]. D'autre part, en plus d'occuper un rôle clinique, la TEP est aussi largement utilisée en recherche.

Les petits animaux, plus particulièrement les souris transgéniques, ont toujours eu un rôle actif dans la recherche. En effet, ceux-ci constituent des modèles idéaux grâce à leur faible coût, leur cycle reproducteur rapide et leur disponibilité [9]. De plus, puisque la plupart des gènes humains ont un équivalent dans une souris, de multiples maladies peuvent être évaluées sur ces sujets. Ensuite, la connaissance du génome de la souris permet d'interchanger des gènes et ainsi mieux comprendre les processus génétiques. Suite à ces avantages, il est bien évident que la recherche effectuée sur les petits animaux occupe un rôle important en TEP. De plus, l'ensemble des résultats, connaissances et découvertes réalisés en TEP sur les petits animaux, peuvent, dans la majorité des cas, influencer la recherche sur les humains ou vice-versa. Pour ces raisons, il existe actuellement un effort important afin de développer des scanners TEP spécifiquement dédiés aux petits animaux.

L'observation de détails infimes sur une souris exige bien évidemment une précision supérieure à la TEP clinique. En effet, pour atteindre le même niveau de détails (environ 3-4 mm) que chez un humain, la résolution spatiale d'un appareil TEP destiné aux petits animaux doit être inférieure à 1 mm [37]. En vue d'atteindre cet objectif, la résolution spatiale des différents appareils s'est constamment améliorée et se rapproche de plus en plus de la précision souhaitée. Cependant, pour conserver le même rapport de signal sur bruit, l'amélioration par un facteur de 10 de la résolution spatiale implique d'augmenter la sensibilité par un facteur de 10^3 [82]. Sans cette sensibilité, le bruit sur l'image devra être filtré, ce qui dégrade considérablement les efforts apportés à la résolution spatiale [9]. Par conséquent, tant qu'un compromis entre ces deux critères de performances antagonistes n'est pas obtenu, l'amélioration de la résolution spatiale n'assure pas une meilleure image tant que la sensibilité n'est pas elle aussi améliorée.

Plusieurs méthodes alternatives ont été envisagées afin d'augmenter le nombre de coïncidences intégrées à l'image et ainsi compenser la lacune en sensibilité des scanners TEP. Entre autres, certaines approches tentent de récupérer des coïncidences habituellement considérées inutilisables, les coïncidences triples. Celles-ci contiennent un troisième événe-

ment radioactif qui est le résultat d'un phénomène physique nommé la diffusion Compton. Sans analyse, cet événement indésirable limite la bonne localisation de la coïncidence au sein de la distribution radioactive et celle-ci ne peut donc pas être intégrée à l'image. Bien que des équations théoriques existent, celles-ci ne sont pas applicables en pratique en raison de l'incertitude des mesures des appareils TEP. D'autres techniques tiennent compte de ces incertitudes, mais sont trop peu performantes pour supporter en temps réel le flux de coïncidences d'un scanner TEP.

En guise d'alternative aux méthodes plus traditionnelles, des recherches antérieures ont utilisé une technique d'intelligence artificielle particulièrement efficace dans des problèmes de classification : les réseaux de neurones artificiels (RNA) [52–55]. Dans un traitement hors ligne, le RNA a été en mesure de classifier correctement 74.7% des coïncidences triples, permettant ainsi d'améliorer la sensibilité de 54.2% au coût d'une dégradation du rapport de contraste sur bruit d'environ 20%. La faisabilité en temps réel n'est cependant pas garantie, car un scanner TEP dispose généralement de peu de ressources de calculs supplémentaires pour intégrer une technique devant faire face à flux de données très important.

1.2 Définition et objectifs du projet de recherche

Dans l'intention d'améliorer la sensibilité d'un scanner TEP qui favorise la résolution spatiale, la récupération des coïncidences triples par un réseau de neurones artificiels semble une technique viable en considérant ses résultats théoriques prometteurs. Le projet de recherche décrit dans ce document tente donc de répondre à la question suivante :

*Comment est-ce qu'un réseau de neurones artificiels peut être en mesure d'effectuer une classification en **temps réel** des coïncidences triples d'un scanner de tomographie d'émission par positrons, malgré les **ressources limitées** de ce dernier ?*

Puisque le contexte du projet se rattache au scanner LabPETTM, l'objectif principal est de concevoir et d'optimiser sur un circuit FPGA un réseau de neurones artificiels avec les ressources limitées du scanner LabPETTM, de façon à ce qu'il soit suffisamment performant pour être capable de classifier en temps réel le flux coïncidences triples provenant de ce dernier.

1.3 Contributions originales

La contribution originale principale de ce projet est de confirmer la faisabilité en temps réel de la technique de récupération des coïncidences triples par réseau de neurones artificiels, et ce, malgré les ressources limitées d'un scanner TEP. Dans l'affirmative, l'approche de réseau de neurones telle que développée par les recherches antérieures [52–55] offre l'opportunité d'être une alternative concrète pour améliorer la sensibilité. Dans une autre mesure, la réussite du projet de recherche contribue à faire connaître la technique des réseaux de neurones dans domaine de la tomographie d'émission par positrons. En effet, cette approche est une approche pertinente à considérer en guise d'alternative aux solutions traditionnelles. Par exemple, la technique des réseaux de neurones artificiels pourrait être utilisée pour s'attaquer au problème des coïncidences fortuites. De plus, puisque l'équipe en place travaille actuellement sur la conception du scanner LabPETTMII, le réseau de neurones conçu pourrait probablement être adapté afin de répondre à un flux de données encore plus important.

1.4 Plan du document

Le **chapitre 2**, intitulé *Tomographie d'émission par positrons*, traite des principes de bases, des critères d'évaluation et des récentes tendances de la TEP.

Le **chapitre 3**, intitulé *Amélioration de la sensibilité par récupération des coïncidences triples*, traite des différentes approches possibles permettant d'analyser et de récupérer les coïncidences triples.

Le **chapitre 4**, intitulé *Réalisation d'un RNA au sein d'un FPGA*, traite des considérations et choix technologiques possibles lors de la réalisation d'un RNA au sein d'un FPGA.

Le **chapitre 5**, intitulé *Intégration au scanner LabPETTM*, décrit l'ensemble des contraintes de performances et de ressources à considérer pour pouvoir intégrer les technique de récupération des coïncidences triples au sein du scanner LabPETTM.

Le **chapitre 6**, intitulé *Conception de la récupération des triples*, décrit l'architecture globale du module conçu exécutant la classification en temps réel des coïncidences triples.

Le **chapitre 7**, intitulé *Méthodologie*, décrit les procédures permettant la reconstruction et l'évaluation d'image avec ou sans récupération des coïncidences triples.

Le **chapitre 8**, intitulé *Résultats*, décrit les résultats obtenus en temps réel suite à l'application de la méthodologie du chapitre précédent.

Le **chapitre 9**, intitulé *Discussion*, discute de l'architecture du module, des méthodes d'acquisition et de résultats obtenus comparativement aux recherches antérieures.

Le **chapitre 10**, intitulé *Conclusion*, est la synthèse du présent document.

CHAPITRE 2

Tomographie d'émission par positrons

2.1 Description générale

2.1.1 Principes de base

La TEP est une technique d'imagerie nucléaire permettant de visualiser de façon non invasive la distribution de sondes moléculaires dans n'importe quelle section du corps [68]. Ces sondes moléculaires, aussi appelées radiotraceurs, se composent d'une molécule comme le glucose auquel un atome a été substitué par un radioisotope typique de la TEP (^{11}C , ^{13}N , ^{15}O , ^{18}F) [32]. Cette substitution au sein de la molécule rend son noyau instable, provoquant à la longue une désintégration radioactive par l'émission d'un positron, transformant le noyau en un état plus stable [11]. Après la perte de son énergie et un court parcours inférieur à 1 mm (0.54 mm RMS pour le ^{18}F) au sein de la matière environnante, le positron émis se combine ensuite avec un électron du milieu dans une réaction d'annihilation [76]. La masse des deux particules est convertie en énergie, apparaissant sous la forme de deux photons d'annihilation, émis à des directions pratiquement opposées d'énergie de 511 keV [81] (figure 2.1).

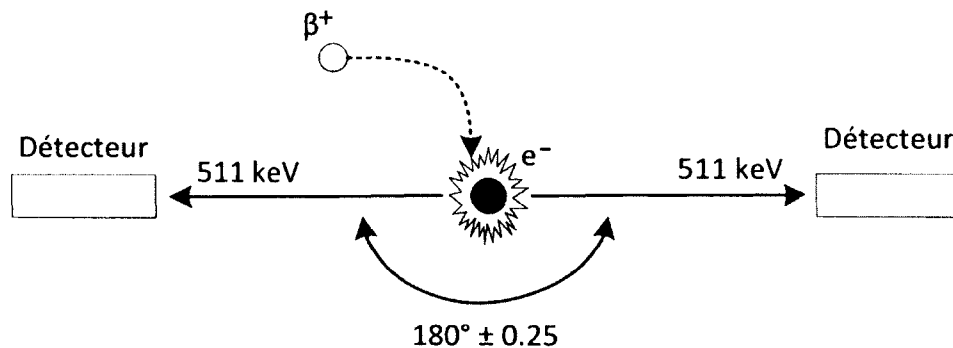


Figure 2.1 L'émission d'un positron et sa réaction d'annihilation.

Afin d'estimer l'origine de l'émission du positron, un scanner TEP placé autour du sujet doit intercepter les deux photons d'annihilation. Pour ce faire, la plupart des détecteurs en TEP sont composés de blocs de cristaux scintillateurs inorganiques très denses couplés à un ou plusieurs photodétecteurs [35]. Les cristaux scintillateurs, efficaces pour arrêter les photons incidents et ainsi absorber leur énergie, génèrent alors de la lumière. Celle-ci est captée par un photodétecteur qui la convertit en un signal électrique dont l'amplitude est proportionnelle à l'énergie initialement déposée par le photon d'annihilation. Un traitement numérique de ce signal permet ensuite de caractériser chaque détection en termes d'énergie et de temps. Lors de la détection quasi simultanée de deux photons d'énergie suffisante au sein d'une même fenêtre temporelle (typiquement de 0,5 à 10 nanosecondes), les deux événements sont alors considérés physiquement corrélés à la même réaction d'annihilation et forment alors une coïncidence [13]. Finalement, malgré que le lieu exact de la réaction d'annihilation demeure inconnu, l'accumulation de millions de coïncidences permettra de reconstruire une image décrivant la distribution spatiale du radiotracer en fonction du temps [2].

2.1.2 Interactions des photons d'annihilation avec la matière

Dans la section précédente, il a été mentionné qu'un traitement numérique du signal à la sortie du détecteur permet de déterminer l'énergie déposée d'un photon d'annihilation. Cependant, bien que cette énergie soit toujours de 511 keV, les mesures imparfaites des détecteurs peuvent limiter la détection totale de l'énergie du photon incident. Ceci est relié au fait que les photons incidents peuvent interagir de deux façons possibles lorsqu'ils entrent en collision avec un atome de la matière qu'ils traversent [44]. Tout d'abord, l'effet photoélectrique représente le cas idéal où la totalité de l'énergie du photon d'annihilation est absorbée par un électron d'une couche atomique interne. L'électron est alors éjecté de l'atome avec une énergie cinétique de 511 keV moins son énergie de liaison et le photon disparaît. La diffusion Compton consiste quant à elle à la deuxième interaction possible entre un photon et un atome. Cette fois-ci, le photon ne transfère seulement qu'une partie de son énergie à un électron d'une couche atomique externe, en plus d'être dévié d'un angle θ qui dépend de l'énergie transférée. Malgré son énergie cinétique moindre, le photon dévié peut ensuite entrer de nouveau en collision avec d'autres atomes, que ce soit par effet photoélectrique ou par une autre diffusion Compton. Finalement, en plus de varier selon l'énergie du photon, la probabilité de l'effet photoélectrique dépend de la composition de l'atome de collision. Celle-ci est donc proportionnelle à ρZ_{eff}^4 où ρ représente la densité et Z_{eff} le numéro atomique [78].

2.1.3 Types de coïncidences

La description précédente des coïncidences a jusqu'à maintenant décrit le scénario idéal où il est présumé que la paire de photons détectés provient de la même réaction d'annihilation et que cette réaction a lieu dans le volume couvert par le détecteur. Ce genre d'événements est considéré comme une coïncidence vraie si en plus des suppositions précédentes, aucun des deux photons d'annihilation n'a été dévié par diffusion Compton préalablement à sa détection [60]. Un exemple est illustré à la figure 2.2A.

Une coïncidence fortuite se produit lorsque deux photons ne provenant pas du même événement d'annihilation sont détectés dans la même fenêtre de temps. Puisque le nombre de détections de photons est proportionnel à la dose de radioactivité injectée, le nombre de coïncidences aléatoires fortuites augmente alors au carré du nombre d'isotopes présents dans la région d'intérêt [60]. Par leur distribution uniforme dans la région d'intérêt, ce genre de coïncidences forme une sorte d'arrière-plan sur l'image reconstruite, ce qui contribue à la dégradation du contraste et du rapport de signal sur bruit [63]. Par exemple, même un rapport modeste de 10% de ce genre d'événement par rapport aux coïncidences vraies provoque l'apparition de distorsion sur les informations quantitatives [28]. L'application d'une compensation soustrayant le nombre approximatif de coïncidences fortuites est donc nécessaire avant d'effectuer des mesures quantitatives. Un exemple de coïncidence fortuite est illustré à la figure 2.2B.

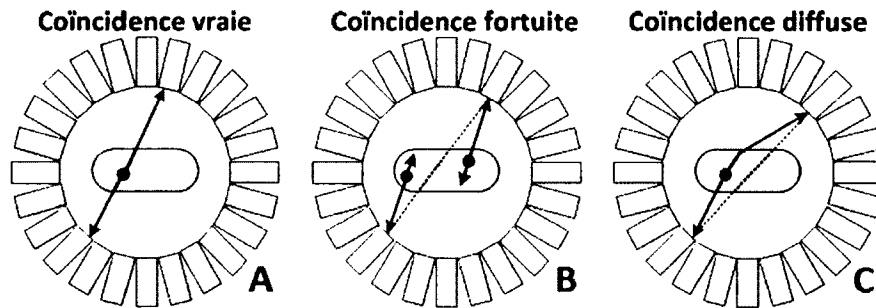


Figure 2.2 Types de coïncidences en TEP.

Une coïncidence diffuse a lieu lorsqu'au moins un des deux photons d'annihilation subit une diffusion Compton ne permettant pas, par la suite, d'en retirer la *vraie* ligne de réponse [11]. L'exemple de la figure 2.2C démontre un exemple de coïncidence diffuse où un des deux photons a subi une diffusion Compton dans le sujet représenté par un ovale, mais il est aussi possible que l'origine de la diffusion Compton soit dans un détecteur. Pour chacune des diffusions Compton subies, le photon est dévié de sa trajectoire originale et est susceptible d'être détecté ailleurs, causant alors l'apparition de coïncidences multiples.

Puisqu'il n'est pas possible de discriminer la totalité de ces événements avec la fenêtre d'énergie relativement large (350-650 keV) [38], ce genre de coïncidence forme lui aussi une source de bruit d'arrière-plan sur l'image.

2.2 Défis de la TEP destinée aux petits animaux

L'observation de détails infimes sur une souris comparativement à un humain exige bien évidemment une précision supérieure. En effet, pour atteindre le même niveau de détails (environ 3-4 mm ou 1 ml) que chez un humain, la résolution spatiale d'un appareil TEP destiné aux petits animaux doit être inférieure à 1 mm (1 μ l en volume) [37]. Un tel degré de précision n'assure cependant pas une meilleure image, car bien que plus petit, chaque voxel de l'image devra conserver un contraste similaire dans un plus grand nombre de pixels. Effectivement, pour conserver le même rapport de signal sur bruit, l'amélioration par un facteur de 10 de la résolution spatiale implique d'augmenter la sensibilité par un facteur de 10^3 [82]. Sans cette sensibilité, le bruit sur l'image devra être filtré, ce qui dégrade considérablement les efforts apportés à la résolution spatiale [9]. Pour ces raisons, la résolution spatiale et la sensibilité sont deux critères de performance cruciaux à optimiser pour la TEP sur les petits animaux.

2.3 Résolution spatiale

La résolution spatiale d'un scanner TEP influence le niveau de détail observable sur une image ainsi que la précision avec laquelle la concentration des radiotraceurs est quantifiée [30]. C'est donc un aspect particulièrement important à considérer, bien évidemment pour l'imagerie de petits animaux, mais aussi pour l'imagerie de petites structures anatomiques chez l'humain. Ce critère est évalué en utilisant l'équation 2.1 [16] :

$$\text{Largeur à mi-hauteur (FWHM)} = \alpha \sqrt{(0,0022D)^2 + r^2 + (d/2)^2 + b^2} \quad (2.1)$$

où α est un facteur relié à la reconstruction de l'image ($1,1 < \alpha < 1,3$), 0,0022 l'incertitude liée à la non colinéarité de la réaction d'annihilation, D le diamètre du scanner, r le déplacement moyen du positron, d la taille des détecteurs et finalement b un facteur lié au décodage effectué dans les détecteurs utilisés. Les effets des paramètres D , r et d sur la résolution spatiale sont d'ailleurs présentés dans les prochaines sections.

2.3.1 Facteurs physiques

Puisque le positron n'a pas nécessairement perdu la totalité de son énergie cinétique lors de la réaction d'annihilation, une variation possible d'environ $0,25^\circ$ peut s'ajouter à l'angle entre les deux photons rendant cet angle non colinéaire [74]. Selon l'équation 2.1, la dégradation de la résolution spatiale pour ce phénomène est directement liée au diamètre D du scanner. Un petit diamètre limiterait donc les effets de la non-colinéarité des photons. Cependant, ce paramètre n'est pas vraiment ajustable puisque le diamètre du scanner est généralement adapté aux dimensions du sujet étudié. Ensuite, l'autre facteur physique non ajustable de l'équation correspond au déplacement moyen du positron avant sa réaction d'annihilation. Puisque ce déplacement dépend de l'énergie impliquée lors de son émission, la dégradation de la résolution variera selon le radionucléide utilisé [46]. Par exemple, le déplacement r inférieur à 0,5 mm du ^{18}F constitue l'une des principales raisons pour son utilisation fréquente.

2.3.2 Dimension des détecteurs

Dans la plupart des scanners TEP, la dimension des détecteurs demeure le facteur prépondérant dans la détermination de la résolution spatiale. En examinant l'équation 2.1, plus le paramètre d est petit meilleure sera la résolution spatiale. Ce constat est simple à valider en prenant pour exemple la figure 2.3A. Sur cette représentation, la diminution de la surface d'interaction du détecteur permettrait de réduire l'incertitude quant à l'origine de l'émission du positron, car le tube de réponse (zone grise) serait plus mince. Les détecteurs minimisant leur distribution d'énergie obtiendraient donc les meilleurs résultats en terme de résolution spatiale [72]. Pour pousser cette analyse plus loin, il est intéressant de considérer la dégradation de la résolution spatiale en dehors du centre du scanner. En effet, lorsqu'un scanner TEP dispose ses détecteurs selon une géométrie circulaire, les photons d'annihilation traversant sans collision un détecteur peuvent être détectés avec un angle d'incidence anormal dans un autre détecteur adjacent [27]. Ce phénomène, l'effet parallaxe, ajoute alors une incertitude quant à l'origine de la réaction d'annihilation, ce qui cause des erreurs de positionnement. Cette incertitude est représentée par l'élargissement du tube de réponse (zone grise) à la figure 2.3B. La dégradation de la résolution spatiale par l'effet parallaxe peut être limitée en utilisant des détecteurs plus petits ou en élargissant le diamètre du scanner pour concentrer le sujet au centre de l'appareil. Cependant, en plus d'augmenter les coûts de l'appareil en nécessitant plus de détecteurs, l'élargissement

du diamètre est contraire à ce qui est préconisé pour la non-colinéarité des photons. Le diamètre doit donc être un compromis entre toutes ces considérations.

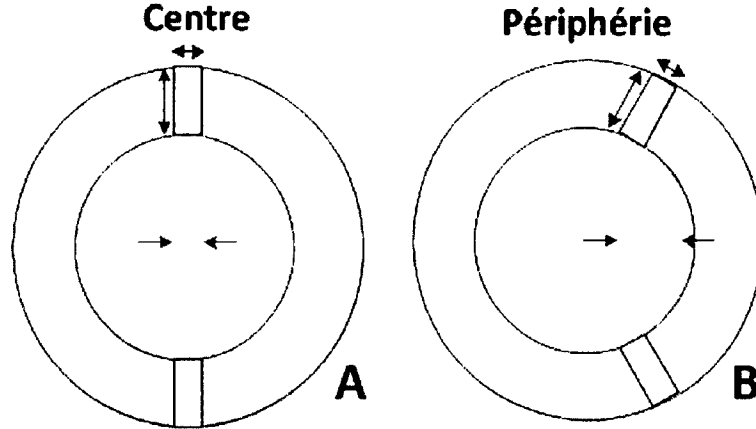


Figure 2.3 Effet parallaxe en périphérie du scanner.

La minimisation des dimensions des détecteurs semble donc une approche plus viable pour améliorer la résolution spatiale. Cependant, cette démarche n'est pas sans limitations. Tout d'abord, un petit détecteur peut être complexe à manufacturer et être très coûteux [32]. Ensuite, la multiplication du nombre de détecteurs pour couvrir le même angle solide complique considérablement l'électronique nécessaire pour intégrer tous les canaux [40]. L'électronique supplémentaire rend d'ailleurs plus difficile de conserver la dissipation de chaleur à un niveau acceptable. De plus, le gain obtenu en résolution spatiale par une miniaturisation des détecteurs se fait généralement au détriment de la sensibilité de l'appareil [46]. En effet, comme il sera discuté plus tard, la taille du détecteur demeure un facteur important dans sa capacité à arrêter un photon et ainsi générer un signal. Le nombre d'événements comptés pourrait ainsi être réduit.

2.3.3 Localisation des détecteurs

Une quantification de la résolution spatiale radiale en fonction de la distance du centre a démontré que la connaissance de la profondeur d'interaction du photon dans le détecteur permettait de réduire les conséquences de l'effet parallaxe [50]. Par exemple, la dégradation était réduite de 45% en sachant dans quelle moitié du détecteur l'interaction du photon a eu lieu. L'utilisation de détecteurs fournissant une information quant à la profondeur d'interaction est alors un choix judicieux permettant d'améliorer la résolution spatiale.

2.4 Sensibilité

La sensibilité d'un appareil TEP, qui représente l'efficacité avec laquelle ce dernier détecte une coïncidence lorsqu'un radionucléide dans la région d'intérêt émet un positron, est un critère important à maximiser pour réduire l'incertitude sur l'ensemble des mesures en temps et en énergie du scanner. Puisque ce critère n'atteint seulement que quelques pourcents dans l'imagerie des petits animaux dans l'imagerie du corps complet d'un humain [40], sa maximisation est d'autant plus importante pour combler l'amincissement des voxels en vue d'améliorer la résolution spatiale. La sensibilité peut donc être évaluée selon l'efficacité de détection d'une source disposée au centre du scanner ou par le taux de comptage obtenu en fonction concentration radioactive [37]. Cependant, aucune de ces techniques ne permet d'évaluer l'effet du bruit d'arrière-plan, causé entre autres par des coïncidences fortuites. Par exemple, la qualité de l'image d'un système récoltant sans bruit 1 million d'événements est susceptible d'être bien supérieure à celle d'un autre système en récoltant 1,5 million dont le tiers des événements sont du bruit d'arrière-plan [76]. La méthode *Noise Equivalent Count* (NEC) [73] permet alors de prendre en considération l'effet du bruit d'arrière-plan sur la sensibilité.

2.4.1 Facteurs de sensibilité

La sensibilité d'un scanner TEP dépend principalement de son efficacité géométrique ainsi que de son efficacité intrinsèque de détection. L'efficacité géométrique correspond à l'angle solide de couverture des détecteurs par rapport à la source radioactive et dépend donc de la grosseur des détecteurs, de la distance de ceux-ci par rapport à la source, ainsi que de leur nombre [8]. Ce facteur géométrique s'optimise donc en couvrant le plus d'espace possible autour du sujet. L'efficacité intrinsèque de détection réfère quant à elle à l'efficacité avec laquelle un détecteur peut absorber un photon incident pour ensuite produire un signal utile [11]. Celle-ci est améliorée en utilisant des détecteurs plus denses et plus longs pour accentuer son pouvoir d'arrêt des photons de 511 keV [51].

2.4.2 Temps mort

Le temps mort est le temps nécessaire à un système pour traiter un événement individuel. Pendant ce laps de temps, tout nouvel événement sur le même détecteur ne peut être mesuré. Dans les systèmes de type non paralysable, tout nouvel événement n'affecte pas le signal généré par le premier et est tout simplement ignoré. Dans un système de type paralysable, ce qui est le cas dans la plupart des scanners TEP, un événement supplémen-

taire introduit un nouveau temps mort pouvant chevaucher le premier, risquant ainsi le rejet de l'ensemble des événements. Pour limiter cet effet et ainsi améliorer sa sensibilité, un appareil TEP se doit donc d'optimiser son temps de traitement. En général, l'utilisation d'un cristal à constante de dégradation rapide permet de réduire le temps mort paralysable [17]. De plus, l'utilisation de petits détecteurs permet de limiter la probabilité d'événements consécutifs dans un même détecteur.

2.5 Contraste

De façon générale, le contraste d'une image décrit la différence d'intensité lumineuse entre deux points. Appliqué à la TEP, ce critère de performance correspond donc à la variation de l'intensité radioactive d'un tissu anormal relativement à celle d'un tissu normal. C'est alors un aspect important à considérer, car cette disparité peut par exemple faciliter la formulation d'un diagnostic. Un appareil TEP doit alors posséder une bonne sensibilité, car ce critère permet de limiter la dégradation du contraste de l'image causée par les variations statistiques des mesures, les effets d'arrière-plan (coïncidences fortuites et diffuses) ainsi que les mouvements du sujet. En effet, comme mentionné précédemment, plus la sensibilité est élevée plus l'incertitude sur les mesures est réduite. De plus, le contraste est amélioré en optimisant la résolution en temps ainsi que la résolution en énergie d'un appareil TEP. Effectivement, l'amélioration de ces deux critères permet de réduire l'erreur associée aux mesures des détecteurs et ainsi ultimement limiter l'incertitude sur la localisation de la réaction d'annihilation.

2.5.1 Résolution en temps

La résolution en temps d'un appareil TEP relève de la précision avec laquelle ce dernier caractérise en temps un événement. Ce critère est important puisqu'il permet d'une part de réduire le nombre de coïncidences fortuites en utilisant une fenêtre temporelle plus étroite, d'autre part de réduire le temps mort et finalement de permettre d'utiliser dans les scanners cliniques un algorithme de reconstruction d'image basé sur le temps de vol des photons [47]. Le placement d'une source ponctuelle radioactive entre deux détecteurs permet d'évaluer l'ensemble des variations statistiques des mesures de temps [13]. La résolution en temps entre ces deux détecteurs correspond donc à la largeur à mi-hauteur (FWHM) de la distribution obtenue. Appliquée à un scanner complet, la résolution en temps correspond à la somme quadratique de toutes les résolutions en temps individuelles provenant chacune de l'ensemble des combinaisons possibles de détecteurs d'un scanner.

2.5.2 Résolution en énergie

La résolution en énergie d'un appareil TEP est une bonne indication quant à sa capacité de distinguer les photons de 511 keV et les photons déviés par diffusion Compton avant détection [30]. Une meilleure résolution en énergie permet donc d'utiliser une fenêtre en énergie mieux définie, réduisant ainsi les coïncidences diffuses et fortuites, sans affecter le nombre de vraies coïncidences [45]. À l'aide du spectre en énergie du système, celle-ci est définie comme le pourcentage entre la largeur à mi-hauteur (FWHM) du sommet photopie et l'amplitude de ce sommet [10]. Au moment de la conception des détecteurs, le choix de cristaux scintillateurs favorisant l'effet photoélectrique et présentant un bon rendement lumineux permettra d'accroître le rapport de photopie [17]. Les cristaux scintillateurs présentent cependant des variations statistiques dégradant la résolution en énergie. En effet, le rendement lumineux des cristaux est un processus non linéaire variant en fonction de l'énergie du photon incident [77], mais aussi en fonction de la profondeur d'interaction de ce dernier [19]. Finalement, la résolution en énergie est aussi affectée par bien d'autres facteurs, comme le bruit ajouté de la chaîne d'acquisition.

2.6 Tendances des dernières années

Tableau 2.1 Performances scanners TEP pour petits animaux

Scanner	FWHM ¹ (mm)	Sensibilité absolue (%)	Fenêtre d'énergie (keV)	Réf.
RAT-PET	2,40	4,30	[250-850]	[7]
Sherbrooke PET	2,10	0,40	> 350	[39]
Quad-HIDAC	1,10	1,90	> 200	[56]
MicroPET	1,80	0,56	> 250	[10]
MicroPET R4	5,10	4,00	> 350	[33]
MicroPET Focus 120/220	1,30	6,50	[250-650]	[36]
Yap-PET	1,80	6,50	> 50	[18]
GE Explorer Vista	1,60	1,70	[250-700]	[36]
ClearPET	1,48	3,80	[250-750]	[25, 90]
Mosaic	2,20	1,30	> 400	[36]
LabPET4	1,28 / 1,42 ²	1,10	[250-650]	[5]
LabPET8	1,28 / 1,42 ²	2,10	[250-650]	[5]
Invcon	1,80	9,30	[250-625]	[4]

¹ Résolution spatiale au centre

² Tangentielle/radiale

Tel que mentionné à la section 2.2, le défi de la TEP consiste à atteindre une résolution spatiale inférieure à 1 mm tout en ayant une sensibilité pouvant supporter cette précision. Le tableau 2.1 présente les critères de performances de plusieurs scanners TEP destinés aux petits animaux. Il est possible d'y constater une certaine tendance où la résolution spatiale s'améliore et où la sensibilité absolue ne semble pas évoluer au même rythme.

En conclusion, bien que la résolution spatiale des scanners destinés aux petits animaux atteigne pratiquement le niveau de détails souhaité, celle-ci ne garantit pas une meilleure image tant que la sensibilité de l'appareil n'est pas elle aussi améliorée.

CHAPITRE 3

Amélioration de la sensibilité d'un scanner

Puisque la sensibilité n'atteint toujours pas les standards souhaités, plusieurs méthodes alternatives ont été envisagées afin de compenser cette lacune des scanners TEP. Tout d'abord, l'allongement du temps d'acquisition est une approche simple à mettre en place, mais ce temps additionnel dégrade la rentabilité du scanner et n'est pas toujours possible selon le temps de demi-vie du radiotracer utilisé. Ensuite, l'augmentation de la dose radioactive injectée permettrait d'augmenter le nombre de comptes par voxel, mais peut être limitée par des considérations pharmaceutiques [29]. De plus, une radioactivité plus intense signifie l'augmentation au carré des coïncidences fortuites, ce qui accentue le bruit sur l'image et peut saturer les détecteurs. D'autres méthodes consistent à améliorer l'efficacité géométrique du scanner en prévoyant le plus petit diamètre possible ou en allongeant sa longueur, mais ces solutions ne sont pas toujours abordables pour des questions de faisabilité et de coûts. Par la suite, afin de récupérer des coïncidences vraies où par exemple un des deux photons d'annihilation n'aurait déposé que partiellement son énergie, une autre pratique consiste à augmenter la tolérance des événements diffus en diminuant le seuil d'énergie [34]. Cette permissivité requiert cependant une utilisation accrue des techniques de compensation et de correction des événements indésirables. Parmi celles-ci, certaines approches s'intéressent aux coïncidences diffuses multiples en tentant de déterminer la vraie ligne de réponse de façon à utiliser ces événements dans la reconstruction de l'image. Plus particulièrement, la plupart des méthodes se concentrent sur le cas le plus probable et le plus simple à analyser, c'est-à-dire les coïncidences triples.

3.1 Équations de Klein-Nishina

En 1928, les chercheurs Klein et Nishina élaborèrent une équation permettant de décrire la probabilité d'interaction des photons à haute fréquence (rayon x et rayon gamma). La dérivation de cette équation permet entre autres d'en déduire la relation entre l'angle et l'énergie lorsqu'une particule subit la diffusion Compton [21]. Par conséquent, il est théoriquement possible à l'aide de cette relation de déterminer la vraie ligne de réponse d'une coïncidence triple. En pratique, en prenant compte l'incertitude apportée par la résolution en énergie typique d'un appareil TEP, qui est typiquement de 25%, l'exactitude des équations n'est pas garantie. En effet, l'étendue résultante des angles de diffusion

possibles d'un photon ne permettra pas dans bien des cas de déterminer avec certitude la séquence des interaction d'un photon.

Plusieurs approches sont d'ailleurs basées sur l'équation de Klein et Nishina. Une première consiste à utiliser une dérivation des équations de Klein-Nishina pour obtenir la projection de la déviation des photons à l'aide de données obtenues par une simulation de Monte-Carlo [12]. Bien que les résultats de cette méthode ne soient pas quantifiés, en localisant avec précision le lieu de diffusion, cette technique permettrait théoriquement d'améliorer la sensibilité et ainsi d'améliorer l'image. Cependant, cette méthode est difficilement transposable à d'autres appareils TEP, puisque les résultats sont obtenus en simulant les propriétés d'un détecteur à base de CZT. Effectivement, malgré son excellente résolution en énergie 2,5%, ce type de détecteur est encore peu utilisé puisqu'il présente une faible résolution en temps (environ 10 ns) [20] et qu'il nécessite un très haut voltage d'opération.

D'autres travaux, basés sur un scanner avec des détecteurs de type photodiode avalanche (PDA), ont évalué la pertinence d'utiliser plusieurs techniques basées sur les équations de cinématique Compton pour résoudre le problème des coïncidences triples [66]. En disposant d'une résolution en énergie plus modeste de 15%, mais tout de même supérieure au standard de l'industrie (25%), le meilleur résultat obtenu parmi toutes les méthodes tentées affichait un succès d'identification de 55,8% de la bonne ligne de réponse. Puisque les détecteurs PDA sont une tendance actuelle, ces résultats insatisfaisants permettent de conclure que l'incertitude des mesures d'énergie d'un scanner TEP ne permet pas d'utiliser les équations de Klein-Nishina afin de résoudre efficacement le problème des coïncidences triples.

3.2 Classification bayésienne

En théorie des probabilités, le théorème de Bayes permet d'établir un lien entre les probabilités à posteriori d'un événement \mathbf{X} d'appartenir à une classe \mathbf{C} en fonction des probabilités à priori et des vraisemblances. En d'autres mots, il permet d'exprimer l'évolution du degré d'appartenance à une classe en fonction des faits. Plusieurs systèmes modernes de reconnaissance statistique (parole, image, etc.) se basent d'ailleurs sur ce théorème pour leurs prises de décisions.

Basée sur le théorème de Bayes, la technique *maximum a posteriori* (MAP) a été utilisée afin d'identifier la séquence d'interactions la plus probable d'un photon diffusé pour ainsi pouvoir déterminer la bonne ligne de réponse entre deux détecteurs [64]. Pour un photon ayant eu N interactions, il y a alors $N!$ séquences d'interaction à évaluer.

Le calcul d'une probabilité MAP d'une séquence doit cependant tenir compte de l'imprécision des détecteurs. En effet, le lieu réel de la collision d'un photon avec le détecteur demeure inconnu puisqu'une interaction est toujours localisée à l'électrode la plus proche. Par conséquent, sans changer l'ordre des interactions, une séquence est modifiée 16 384 fois en variant le positionnement de ses interactions, obtenant ainsi 16 384 probabilités différentes. La moyenne de ces probabilités constitue alors la probabilité MAP de la séquence d'interaction.

En ce qui concerne les résultats de cette technique, la classification des coïncidences triples obtient un taux de succès de 76,7%. De plus, même si cette technique est basée sur des simulations d'un scanner avec des détecteurs à base de CZT possédant une résolution en énergie de 2,5%, les auteurs des travaux mentionnent qu'elle est fonctionnelle jusqu'à une résolution de 12,5%. Toutefois, cette solution ne semble pas abordable une fois appliquée à des vraies données de scanners, car le traitement d'un seul événement requiert environ 5 ms au sein d'une carte graphique GeForce 9800 GX2 (NVIDIA) optimisée par la librairie CUDA [59]. En effet, le traitement des dizaines de millions de coïncidences triples retarderait de plusieurs heures ou même plusieurs jours l'obtention d'une image.

3.3 Réseaux de neurones artificiels

Un réseau de neurones artificiel (RNA) est une technique d'intelligence artificielle particulièrement efficace dans des problèmes de reconnaissance de forme et de classification. Par sa capacité à extraire certains des traits distinctifs d'une situation grâce à un apprentissage supervisé, cette méthode est pratique à utiliser puisqu'elle ne nécessite aucune paramétrisation [71]. C'est pourquoi cette technique en remplace souvent d'autres plus traditionnelles dans une multitude de domaines : détection de maladies sur des images médicales [48], détecteur d'échec de capteur [58], localisation et suivi de véhicules [22], etc.

Un neurone artificiel individuel est un classificateur non linéaire. La sortie du neurone est la somme pondérée de ses entrées, suivie d'une fonction d'activation non linéaire [24]. Plus concrètement, face à un nouvel ensemble de n entrées, un neurone artificiel effectue la somme de chaque entrée x_i multipliée par son poids respectif w_i , puis compare cette somme à une valeur de biais b . La sortie y du neurone est ensuite déterminée par une fonction d'activation f , typiquement représentée par une fonction sigmoïde ou tangente hyperbolique. Toutes ces opérations se résument par l'équation 3.1.

$$y = f \left(\sum_{i=1}^n (x_i w_i) + b \right) \quad (3.1)$$

Le regroupement en couches de plusieurs neurones artificiels forme alors un réseau de neurones artificiels qui peut être considéré comme un classificateur non linéaire plus élaboré. Un exemple de neurone artificiel et d'un réseau de neurones est représenté à la figure 3.1.

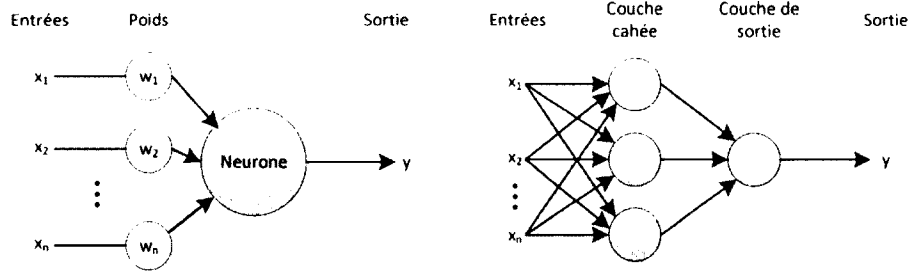


Figure 3.1 Neurone artificiel (gauche) et réseau de neurones artificiels (droite).

L'architecture d'un réseau de neurones artificiels, c'est-à-dire le nombre de couches et le nombre de neurones pour chaque couche, demeure initialement inconnue face à une nouvelle situation. Pour découvrir celle répondant le mieux au problème, l'approche empirique est privilégiée [23, 69]. En effet, face à l'infinité de combinaisons possibles, il n'y a pas nécessairement de solution meilleure que les autres. Cependant, il a été prouvé qu'un réseau n'ayant qu'une seule couche cachée disposant suffisamment de neurones est en mesure d'extraire de façon autonome les traits distinctifs de n'importe quelle situation [15].

Pour être efficace, un RNA doit être entraîné. Effectivement, lorsque soumis qu'à une partie des données d'un problème, un RNA est en mesure de généraliser ses apprentissages pour ensuite pouvoir les appliquer avec succès sur l'autre partie des données [57]. L'objectif de l'entraînement est alors de permettre au RNA de généraliser une situation, autrement dit de lui permettre d'en extraire les traits distinctifs [23]. Plus concrètement, l'entraînement visera à déterminer les poids et les biais de chacun des neurones, c'est-à-dire les paramètres servant à effectuer la moyenne pondérée. Parmi les méthodes d'entraînement, la plus populaire est sans contredit la technique de rétro-propagation de l'erreur développée [69]. Celle-ci consiste à superviser l'entraînement d'un RNA en lui fournissant un ensemble de données à traiter où la vraie réponse est connue. Lors d'une erreur de classification, une correction est propagée à tous les neurones à partir des neurones de sortie, permettant alors à chacun des neurones d'ajuster ses paramètres (poids et biais). À

la longue, le RNA convergera vers une configuration de poids et de biais. Puisque l'entraînement est un processus stochastique, chaque entraînement convergera généralement vers une solution différente des autres et qui n'est pas nécessairement optimale. En critique aux RNA, il n'est pas vraiment possible de comprendre pourquoi une configuration est meilleure qu'une autre, et une incertitude persistera toujours quant à savoir si la solution optimale a été atteinte. Lors de l'entraînement, la convergence des réseaux de neurones artificiels peut parfois être très longue, et par son caractère stochastique, doit être répétée plusieurs fois pour découvrir une solution idéale. L'utilisation d'un moment et d'un taux d'apprentissage adaptatif sont des techniques qui permettent alors d'accélérer grandement la convergence d'un RNA [80].

Afin de résoudre le problème des coïncidences triples, un RNA a été conçu afin de déterminer, à l'aide de la position géométrique et de l'énergie de chacune des détections, quelle est la véritable ligne de réponse d'une coïncidence triple [52–55]. Cependant, puisque la bonne réponse est inconnue dans un vrai scanner TEP, l'entraînement du RNA doit s'effectuer avec des données de simulation. Celles-ci ont donc été obtenues grâce au logiciel GATE [61] par une simulation Monte-Carlo du scanner LabPET4, qui présente une résolution en énergie d'environ 25%.

Avant d'être présentées au RNA, les données en entrées doivent tout d'abord subir un pré-traitement visant à réduire la complexité des données. Plusieurs opérations géométriques sont alors effectuées sur les données initiales et permettent de rendre triviale la position du photon photoélectrique, de regrouper les situations symétriques et d'annuler la composante axiale des données. Malgré un coût additionnel en calcul, cette étape semble être un moindre prix à payer pour réduire la grosseur du réseau de neurones artificiels, puisque par exemple le nombre de dimensions en entrée passe de 12 à 6.

En ce qui a trait aux résultats, le RNA obtient tout d'abord un succès de classification des coïncidences triples de 74,7% lorsqu'il est soumis à un ensemble distinct de données de simulation. Ensuite, lorsqu'il est appliqué à de vraies données provenant du scanner LabPET4, celui-ci permet d'augmenter de 54,2% la sensibilité du scanner, au coût d'une dégradation du ratio de contraste sur bruit. Malgré ces résultats prometteurs, l'auteur [52–55] mentionne que ceux-ci sont obtenus par un traitement hors ligne qui n'est pas optimisé. L'intégration de cette technique dans un vrai scanner en temps réel demeure donc à être prouvée, puisque les ressources nécessaires ainsi que le temps de traitement en découlant sont inconnus.

En conclusion, parmi les méthodes proposées afin d'améliorer la sensibilité via la récupération des coïncidences triples, l'approche des réseaux de neurones artificiels semble l'alternative la plus prometteuse pour effectuer cette tâche en temps réel dans un scanner de tomographie d'émission par positrons.

CHAPITRE 4

Réalisation d'un RNA dans un FPGA

4.1 Faisabilité

Plusieurs applications de réseaux de neurones artificiels fonctionnant dans un FPGA ont été déjà réalisées. Parmi celles-ci, notons la localisation et vérification d'identité des visages sur vidéo à un taux de succès de 92% [88] et la reconnaissance vocale au taux de succès de 96,83% [62]. Ensuite, plusieurs preuves de concept ont aussi été réalisées dans le domaine de la TEP : un estimateur de position de rayon gamma au sein d'un détecteur [1], un réducteur de coïncidences fortuites [67] ou une amélioration des techniques de mesures temporelles [43]. Par constatation de ces exemples, on peut conclure qu'un réseau de neurones artificiels est réalisable sur un FPGA. Cependant, puisque chaque problématique est différente, on ignore les ressources nécessaires pour faire face au problème des coïncidences triples.

4.2 Représentation des données

La représentation des données est un aspect essentiel à évaluer afin d'assurer au système d'atteindre une précision minimale. Tout d'abord, un choix doit être effectué entre une représentation arithmétique en point fixe ou en point flottant. La première représentation est la plus conventionnelle, où la partie entière et la partie fractionnaire sont séparées par une virgule fixe. Cette simplicité lui confère donc des avantages de rapidité, tout en lui permettant de couvrir une plage raisonnable de données. Pour sa part, la représentation en point flottant est définie selon le standard IEEE 754 et est décrite par un signe, un exposant et une mantisse. C'est donc dire que pour un même nombre de bits, une représentation en point flottant des données offre une précision accrue par rapport à la représentation en point fixe, mais à un coût supérieur en ressources [85].

Dans le cas d'une implémentation sur un FPGA, la représentation en point fixe est généralement préférée pour sa simplicité et sa rapidité. Celle-ci comporte cependant le défi de déterminer le nombre minimal de bits pour bien représenter le problème tout en satisfaisant des contraintes de précision. Pour ce faire, plutôt que de tenter l'ensemble des combinaisons possibles, certaines méthodes proposent la technique de recuit simulé pour

optimiser la partie fractionnaire et l'arithmétique affine pour évaluer l'étendue des données [41]. Appliquée aux réseaux de neurones, l'accent de ces techniques doit être mis sur la précision de la partie fractionnaire. Effectivement, puisqu'il est d'usage courant de normaliser les entrées et de ne pas permettre des valeurs très élevées aux poids, il devient facile d'identifier l'étendue des données.

4.3 Représentation des fonctions complexes

La réalisation d'un réseau de neurones ainsi que toutes les étapes préliminaires de simplifications requiert plusieurs opérations complexes pour un FPGA. En effet, l'on retrouve des opérations généralement non natives comme des fonctions trigonométriques, des multiplications, des divisions, des exponentielles, etc. Heureusement, plusieurs techniques permettent de reproduire simplement ces fonctions. Le choix de la bonne méthode s'effectue en fonction du cahier des charges du système, qui permet de déterminer le compromis nécessaire entre performance et minimisation de ressources.

L'utilisation d'une *Look-up table* (LUT) permet d'entreposer dans un bloc mémoire des valeurs discrètes de la fonction à reproduire. Le résultat de la fonction à une entrée s'obtient alors très rapidement en accédant tout simplement à l'espace mémoire approprié. En contrepartie, la mémoire supplémentaire nécessaire peut s'avérer très volumineuse selon la précision souhaitée et de la plage d'entrées à couvrir [6]. Par conséquent, une LUT n'est peut-être pas le bon choix pour des systèmes à mémoire limitée ou exigeant une haute précision.

Une alternative peu coûteuse en mémoire consiste à effectuer l'approximation d'une fonction par des courbes polynomiales de premier ou de second degré. Effectivement, la plupart des approches [42, 79] divisent en régions la fonction désirée et n'ont qu'à conserver dans une mémoire distribuée (registres) les coefficients de chacune des sections. Cette démarche est cependant dépendante à la disponibilité de multiplicateurs matériels, car la création de multiplicateurs logiques peut s'avérer coûteuse en ressources. Finalement, la performance peut être aussi rapide qu'une LUT, mais la complexité requise requiert bien évidemment plus de ressources logiques.

Une autre technique, l'algorithme CORDIC, est simple et efficace pour calculer des fonctions hyperboliques et trigonométriques en utilisant uniquement des additions et des décalages binaires. Effectivement, à l'aide de rotations d'angle arbitraires, l'algorithme converge itérativement vers l'angle de rotation souhaité [3]. Cette approche de calcul possède un

coût modeste en ressources logiques, mais son processus itératif la rend beaucoup plus lente que les deux autres approches.

Finalement, l'arithmétique distribuée est une technique permettant de calculer le résultat d'une somme de multiplications, par exemple un produit vectoriel, sans effectuer de multiplications [83]. En présumant qu'un des deux termes de la multiplication est toujours un coefficient fixe, et que le second terme est variable, mais compris dans l'intervalle $[-1, 1]$, cette méthode emmagasine en mémoire l'ensemble des résultats de multiplications dans des LUT, et permet un calcul rapide des résultats. Comparativement aux autres méthodes, cette technique permet d'atteindre une excellente précision très rapidement, mais elle ne s'applique pas nécessairement à toutes les situations.

4.4 Implémentation d'un réseau de neurones artificiels

En lien avec l'équation 3.1, chaque neurone implémenté doit posséder des blocs de calculs permettant de faire des additions et des multiplications où le nombre nécessaire d'opérations variera en fonction du nombre n d'entrées. Il serait aussi possible d'étudier la possibilité d'utiliser l'arithmétique distribuée telle que décrite précédemment.

La reconstitution exacte de fonctions d'activations comme la fonction sigmoïde ou la tangente hyperbolique s'avère très complexe puisque la plupart des architectures FPGA n'intègrent pas des modules arithmétiques de division, de puissance ou d'exponentielle [70]. Cependant, comme indiqué dans la section précédente, plusieurs alternatives sont disponibles pour représenter cette fonction complexe de façon discrète. D'ailleurs, la discrétisation de la fonction d'activation n'aura pas d'impact négatif sur la performance globale de classification du réseau de neurone complet, pourvu que l'entraînement soit réalisé avec cette même fonction discrétisée [43].

À l'image de la conception d'un seul neurone artificiel, l'exécution complète du réseau de neurones artificiels est un compromis entre performance et ressources logiques. L'implémentation la plus simple ne nécessite qu'un seul neurone artificiel et un engin de contrôle qui séquence les calculs des sorties pour chacun des neurones artificiels. Bien qu'elle nécessite peu de ressources logiques, cette façon de faire ne tire pas avantage du parallélisme d'un FPGA et est, somme toute, peu performante. En effet, en prenant comme exemple une application de reconnaissance vocale, le ratio entre la performance obtenue en fonction des ressources utilisées est 12 fois supérieur lorsque la solution est parallélisée [62].

La parallélisation du réseau de neurones offre une certaine latitude sur les choix de conception, encore une fois orientés en terme de performance par rapport aux ressources logiques. En exécutant parallèlement l'ensemble des couches d'un réseau de neurones [89], un gain en performance considérable peut être obtenu, car le temps d'exécution devient égal à celui de la couche la plus longue à évaluer. Cependant, l'investissement en ressources peut améliorer de façon non significative la performance dans des cas où, par exemple, les temps d'exécution des plus petites couches sont très petits par rapport à la plus large grande. Dans un autre ordre d'idée [26], une différente implémentation en parallèle consiste à uniquement concevoir la couche de neurones la plus large et de disposer d'un engin de contrôle multiplexant les paramètres aux neurones s'exécutant en parallèle. Cette méthode permet de restreindre les ressources nécessaires, plus particulièrement pour les réseaux les plus larges. Par contre, l'engin de contrôle nécessite bien évidemment une charge de calculs supplémentaires.

En conclusion, afin de répondre au cahier des charges du système, l'implémentation d'un réseau de neurones artificiels au sein d'un FPGA correspond à un compromis entre la performance et la minimisation de ressources.

CHAPITRE 5

Intégration au scanner LabPET™

Les travaux de recherche décrits dans ce mémoire se concentrent sur le LabPET™, un scanner TEP à base de photodiodes avalanches conçu à l'Université de Sherbrooke.

5.1 Architecture du LabPET™

Premier de sa catégorie à être commercialisé (Gamma Medica, Inc.), le LabPET peut supporter jusqu'à 4608 détecteurs et est disponible en trois versions se distinguant par leur champ de vision axial respectif, soit 3,75 cm, 7,5 cm et 11,25 cm [5]. Le scanner LabPET™ se base sur des détecteurs phoswich LYSO/LGSO regroupés en groupe de quatre au sein d'un emballage hermétique de Kovar (figure 5.1a). Ceux-ci sont répartis au sein d'anneaux de façon à obtenir au centre du scanner un cercle d'un diamètre de 16.2 cm (figure 5.1b). Les anneaux sont empilés les uns sur les autres et décalés angulairement d'un demi-détecteur (figure 5.1c). Les versions possibles mentionnées précédemment sont donc l'agglomération de 8, 16 ou 24 anneaux, ayant respectivement 1536, 3072 ou 4608 détecteurs.

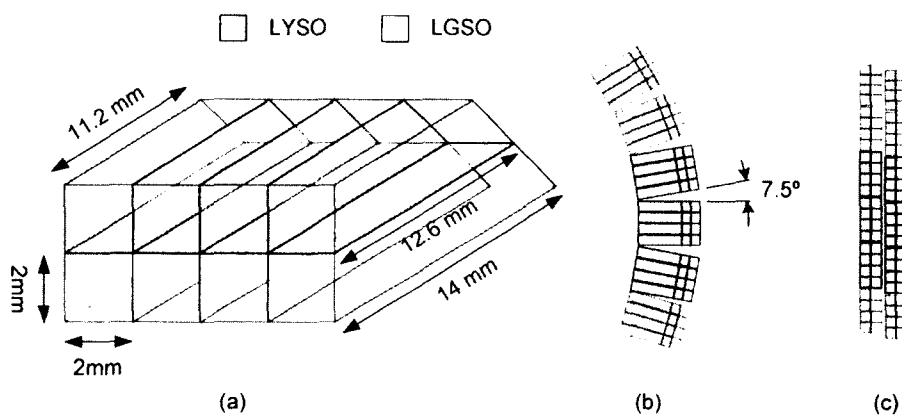


Figure 5.1 Géométrie du LabPET™[53]

Les images des recherches antérieures [53] sur la récupération des coïncidences triples ont été obtenues à partir de la configuration **LabPET4** (3,75 cm). Pour faciliter la comparaison des résultats, cette version est aussi choisie pour les travaux de recherche actuels.

5.2 Acquisition des événements

La caractérisation en temps et en énergie des événements détectés est accomplie par les différentes cartes d'acquisition à 64 canaux. Plus précisément, chaque canal est une chaîne d'acquisition amplifiant le signal à la sortie d'un détecteur et l'échantillonnant à l'aide d'un convertisseur analogique-numérique 8 bits opérant à 45 MHz. Toute détection d'un front est donc vraisemblablement causée par la collision d'un photon d'annihilation au sein d'un détecteur. L'analyse de l'amplitude du signal obtenu permet ainsi de définir l'énergie de l'évènement, car cette amplitude est proportionnelle à l'énergie totale déposée par le photon. Une étampe de temps est alors assignée à l'évènement à l'aide d'une horloge d'une fréquence de 45 MHz. Tout nouvel événement est ensuite acheminé au processeur DSP de la carte d'acquisition. Celui-ci identifie le cristal impliqué de façon à pouvoir appliquer la fenêtre en énergie appropriée et ainsi amoindrir le bruit du système. Les événements respectant le critère d'énergie quittent finalement la carte d'acquisition et sont dirigés en ordre chronologique vers l'engin de coïncidences après avoir subi plusieurs tris successifs. Ceux-ci sont d'ailleurs représentés par un paquet de 8 octets où la répartition des bits est démontrée dans le tableau 5.1.

Tableau 5.1 Représentation du paquet d'un événement [75]

Bits	63	40	39	32	31	24	23	16
Valeur	Temps		Énergie		Baseline		Autres	
Bits	15	8	7	4	3	2	1	0
Valeur	Détecteur		Anneau		Cristal		Autres	

Dans l'état actuel du système, l'énergie contenue dans ce paquet ne peut pas être utilisée dans la technique de récupération des coïncidences triples. En effet, celle-ci ne correspond pas à une valeur exprimée en keV, mais plutôt en une valeur numérisée de l'ADC du canal correspondant. En conséquence, deux valeurs d'énergie ne peuvent donc pas être comparées entre elles sans les histogrammes d'énergie des canaux. Heureusement, cette information est disponible au sein du processeur DSP et permet de convertir l'énergie en keV. Pour ce faire, une simple règle de trois est ajoutée à la fin du code DSP, et ce, sans aucun impact sur le restant du système. Finalement, puisque l'énergie ne peut être encodée que sous 1 octet ou 256 valeurs, la résolution doit être de 3 keV par bit pour bien représenter l'ensemble du spectre en énergie d'un photon d'annihilation.

5.3 Formation des coïncidences

La détection quasi simultanée de deux photons d'énergie suffisante au sein d'une même fenêtre temporelle de temps maximal τ implique que les deux événements sont physiquement corrélés à la même réaction d'annihilation. L'engin de coïncidence, qui reçoit les événements triés en ordre chronologique, regroupe alors ceux qui sont rapprochés temporellement pour ainsi former les coïncidences. Pour ce faire, une fenêtre de temps prend origine au premier événement reçu et est étendue pour tenter d'inclure le prochain événement (figure 5.2). Pour chaque nouvel événement ajouté, cette étape est prolongée jusqu'à ce qu'aucun nouvel événement ne soit trouvé. Lorsqu'un événement se retrouve seul, celui-ci est considéré inutilisable et est automatiquement rejeté. Tout regroupement forme alors une coïncidence qui est ensuite transmise à la collimation électronique.

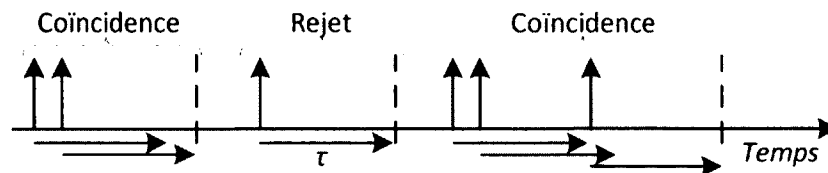


Figure 5.2 Groupement des événements (flèches verticales) en coïncidences [75]

La formation des coïncidences regroupe toutefois des événements non corrélés et génère des coïncidences indésirables dites fortuites. Néanmoins, puisque la distribution statistique des coïncidences fortuites dans la région d'intérêt est uniforme, celle-ci peut être estimée et soustraite pour ainsi limiter sa dégradation du contraste de l'image. En conséquence, l'engin de coïncidence permet aussi d'effectuer cette estimation en temps réel avec l'aide de la technique de la fenêtre décalée [75]. Essentiellement, cette méthode vise à briser la corrélation entre deux événements d'une coïncidence vraie en introduisant une nouvelle séquence d'événements retardés. Ainsi, tous les événements reçus sont dupliques, retardés d'un temps supérieur à τ puis combinés aux événements originaux (figure 5.3). Après un nouveau tri chronologique, la nouvelle chaîne d'événements subit elle aussi le regroupement en coïncidences. Cependant, seules les coïncidences contenant au moins un événement original et un retardé seront retenues. Celles-ci seront alors considérées fortuites et seront elles aussi soumises à la collimation électronique.

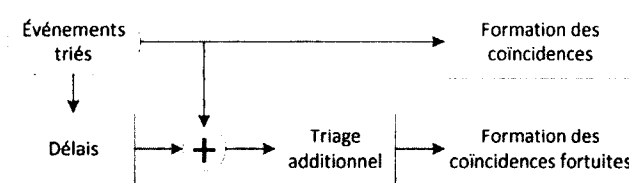


Figure 5.3 Chaînes de formation des coïncidences

5.4 Collimation électronique

La collimation électronique a pour objectif de conserver uniquement les coïncidences pouvant être utilisées dans la reconstruction de l'image, c'est-à-dire toutes les coïncidences dont la ligne de réponse passe dans le champ de vue du scanner. Cette opération consiste donc en une simple validation géométrique appliquée successivement dans le plan radial et le plan axial. La figure 5.4 présente un exemple de collimation électronique radiale où seule la ligne de réponse de la figure 5.4 b est valide.

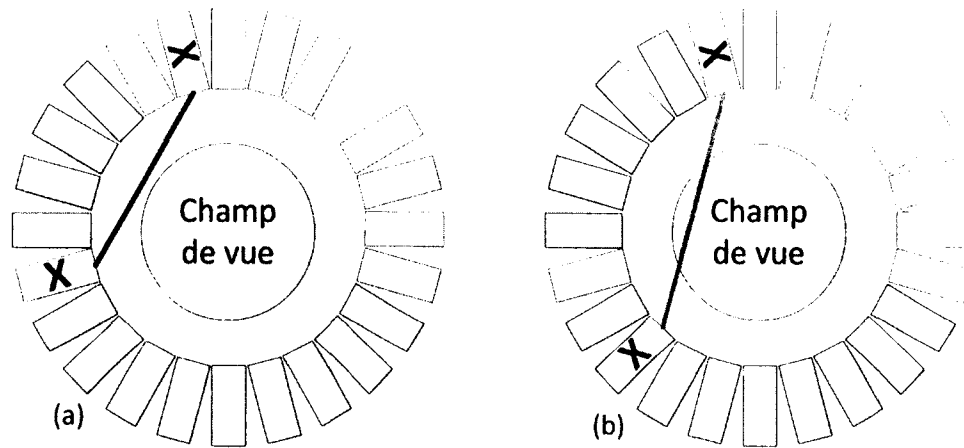


Figure 5.4 Collimation électronique radiale d'une coïncidence double. (a) Coïncidence rejetée et (b) coïncidence conservée.

Au même titre que la formation des coïncidences, la collimation électronique sur les coïncidences fortuites s'effectue dans une chaîne séparée et indépendante (figure 5.5). En revanche, malgré leur séparation, les deux modules effectuent exactement le même traitement.

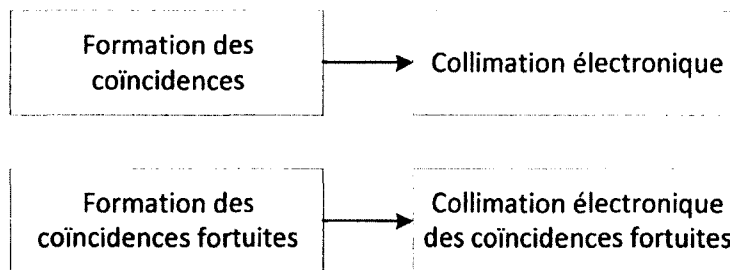


Figure 5.5 Collimation électronique radiale d'une coïncidence double.

Cette validation géométrique est cependant plus problématique à appliquer dans le cas d'une coïncidence multiple, car celles-ci peuvent posséder plus d'une ligne de réponse traversant le champ de vue. En effet, ces coïncidences multiples peuvent être le résultat d'une multitude de scénarios possibles et leur inclusion dans la reconstruction de l'image sans analyse supplémentaire pourrait constituer une source de bruit significative. C'est pourquoi la conception actuelle intègre uniquement des cas comme la figure 5.6a où une seule des lignes de réponse traverse le champ de vue du scanner. Les cas similaires à la figure 5.6b entraînent donc un rejet de la coïncidence.

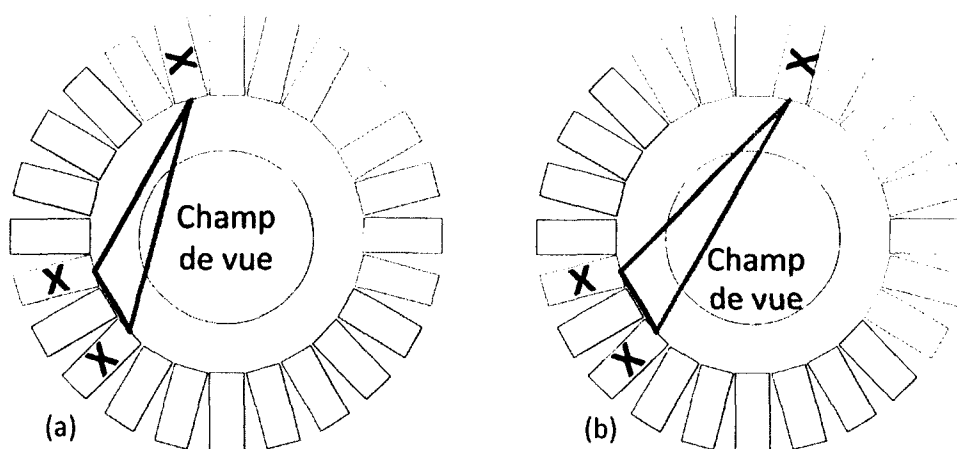


Figure 5.6 Collimation électronique radiale d'une coïncidence triple. (a) Coïncidence triple conservée et (b) coïncidence rejetée.

5.5 Extraction des données

Afin d'extraire les données permettant de reconstruire une image, les coïncidences valides sont d'abord emmagasinées dans une mémoire vive servant de tampon puis sont envoyées vers un ordinateur. Puisque les coïncidences ne sont pas les seules données transférées, un entête précède celles-ci et permet à la reconstruction de l'image d'en distinguer le contenu.

Cette étape est le point d'engorgement du système, car l'écriture sur le disque est relativement lente par rapport aux autres opérations. Par conséquent, le système est donc limité dans sa configuration actuelle à un débit moyen d'écriture d'environ 18 Mo/s. Tout flux de données plus rapide pourrait donc entraîner le débordement de la mémoire vive et la perte de données.

De façon à permettre le contrôle du flux de données ou la taille de ces données, différents modes d'acquisition sont disponibles dans le scanner LabPET4. Tout d'abord, le mode collimation électronique correspond exactement aux étapes présentées jusqu'à maintenant et garantit l'obtention de coïncidences valides géométriquement. Le mode recherche constitue quant à lui à la désactivation de la collimation électronique. Le flux de données devient donc beaucoup plus important, car toutes les coïncidences sont transférées. Finalement, le mode image permet de réduire la taille des données et de ne conserver uniquement que les informations essentielles à la reconstruction de l'image (temps, énergie et identification du détecteur). Ce mode est particulièrement utile lors de l'acquisition de sources radioactives à activité élevée où le débit de données est très important.

5.6 Insertion de la récupération des coïncidences triples

Le module de récupération des coïncidences triples doit être inséré au sein de l'engin de coïncidences après la formation des coïncidences. De plus, puisqu'il faut appliquer cette méthode en temps réel, celle-ci doit assurément être appliquée avant l'extraction des données. D'autre part, étant donné que cette technique tente d'identifier la coïncidence double originale, toute coïncidence double obtenue doit nécessairement être validée géométriquement par la collimation électronique. Le diagramme de séquence présenté à la figure 5.7 respecte ces conditions en détournant les coïncidences triples pour ensuite rediriger les cas considérés valides vers la collimation électronique.

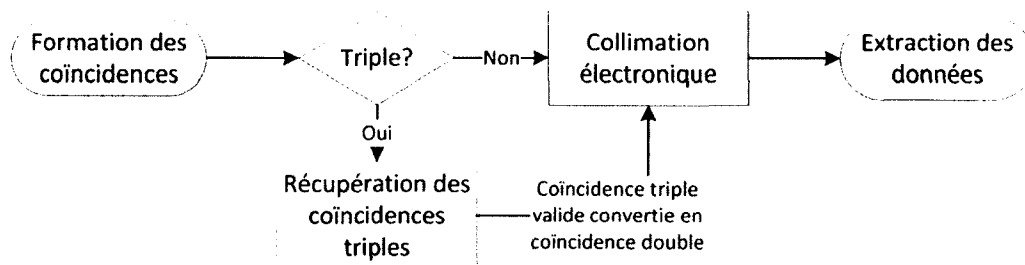


Figure 5.7 Insertion du traitement avant la collimation électronique.

En revanche, malgré sa viabilité et sa simplicité de conception, les résultats obtenus d'une acquisition avec cette architecture seraient difficilement quantifiables. En effet, les coïncidences triples converties en coïncidences doubles ne seraient pas distinguables des coïncidences initialement doubles. De plus, les coïncidences triples originellement acceptées par la collimation électronique, mais rejetées par la récupération des triples seraient perdues. Par conséquent, cette solution ne permettrait pas d'effectuer des images du type

avant-après sur une même acquisition. Dans l'intention de permettre la comparaison des résultats sur une même acquisition, l'ensemble de l'information transitant par le module de récupération des coïncidences triples doit être acheminé à l'extraction des données. En ce sens, la séquence précédemment proposée est modifiée à la figure 5.8 de façon à détourner complètement les coïncidences triples de la collimation électronique.

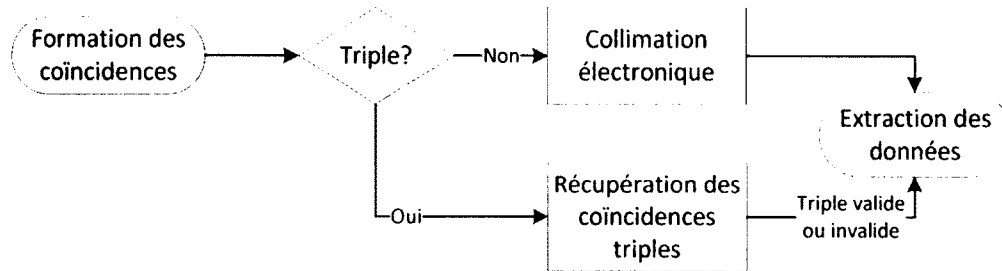


Figure 5.8 Détour complet des coïncidences triples.

L'envoi direct des coïncidence triple à l'extraction des données permet d'ailleurs de modifier le paquet de données envoyé en transmettant les informations suivantes :

- un entête distinct permettant d'identifier une coïncidence triple ;
- le résultat de la récupération des coïncidences triples et
- les informations de base des trois événements (énergie, temps et détecteur).

L'architecture obtenue de l'engin de coïncidence prend alors la forme de figure 5.9, où les nouveaux éléments sont en vert. Par souci d'économie de ressources, la récupération des coïncidences triples n'est pas dupliqué et traite aussi les coïncidences triples de la chaîne d'estimation des fortuits.

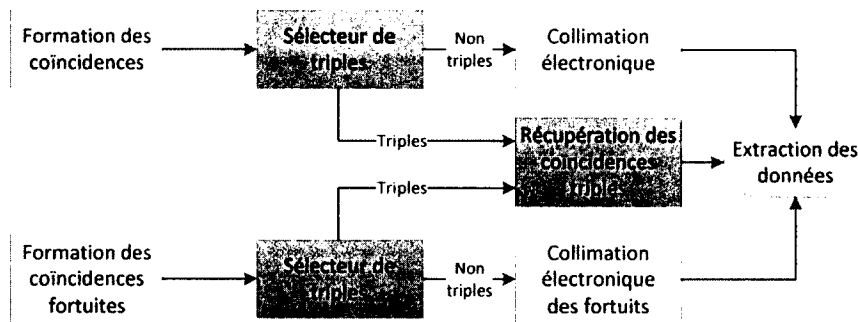


Figure 5.9 Insertion de la récupération des coïncidences triples.

L'application de différents filtres sur le fichier d'acquisition à l'aide d'un traitement hors ligne permet alors de reproduire le traitement originel et le résultat qu'aurait produit

la séquence de la figure 5.7. Cette étape hors ligne ne compromet pas l'évaluation en temps réel de la technique de la récupération des coïncidences triples puisque l'essentiel du traitement est déjà accompli en temps réel. Finalement, puisque ce traitement hors ligne nécessite toutes les informations d'un événement, il est essentiel de ne pas compresser les paquets de données. Le mode d'acquisition ne peut donc pas être en mode image.

5.7 Respect des contraintes

5.7.1 Contraintes de performance

Les travaux antérieurs estiment que le taux maximal de triples dans le scanner LabPET4 s'élève à environ 1 million par seconde à une activité de 180 MBq [53]. Puisque l'engin de coïncidence où sera intégré le module de récupération des coïncidences triples opère à un horloge cadencée à 50 MHz, la récupération d'une coïncidence triple doit donc avoir une durée maximale de 1 μ s ou de 50 cycles d'horloge. De plus, l'ensemble des temps de propagation des signaux doit être inférieur à 20 ns. Dans ces circonstances, il est très difficile voire impossible de concevoir une architecture séquentielle pouvant réaliser l'ensemble des opérations requises en seulement 50 cycles d'horloges. Face à ce constat, il devient donc impératif de tirer avantage du parallélisme des cartes FPGA et de concevoir une architecture en pipeline. Une architecture en pipeline peut être définie comme la décomposition d'un processus répété et séquentiel en plusieurs sous-processus autonomes et parallèles. Chaque sous-processus possède ainsi ses propres ressources dédiées et est coordonné avec les autres par un engin de contrôle global. De cette manière, plusieurs entrées peuvent être traitées en parallèle et ainsi réduire le temps mort du système, même si le traitement global d'une entrée est possiblement plus long.

5.7.2 Contraintes de ressources

L'engin de coïncidence du LabPET4 s'exécute au sein d'un FPGA XC2VP50 [86] dédié (série Virtex 2 Pro). En plus de permettre la synthèse de code VHDL, ce modèle permet l'utilisation de code C compilé dans un processeur matériel PowerPC [31] ou dans un processeur synthétisé MicroBlaze [87]. L'analyse du tableau 5.2 révèle que la contrainte principale en termes de ressources est liée au nombre d'unités logiques disponibles. En effet, contrairement aux autres types de ressources présentées, le nombre d'unités logiques peut varier rapidement et est difficile à prévoir. La conception du module de récupération des coïncidences triples doit alors minimiser le plus possible cet effectif, avec entre autres la possibilité à utiliser les différents processeurs disponibles.

Tableau 5.2 Ressources disponibles de l'engin de coïncidence du LabPET4

Ressource	Total	Utilisés	Disponibles	Taux de disponibilité
Unités logiques (Slices)	23 616	12 340	11 276	48%
Blocs RAM 1 Ko	232	107	125	44%
Multiplicateurs 18 bit	232	0	232	100%
Processeur PowerPC	2	1	1	50%

Le MicroBlaze est un processeur synthétisé RISC 32 bits pouvant exécuter 34 millions d'instructions par seconde lorsque cadencé à 50 MHz. Dans sa version la plus épurée, celui-ci est synthétisé à un coût d'environ 500 unités logiques par processeur. Par son coût élevé et sa performance insuffisante pour le contexte de l'application, le processeur MicroBlaze n'est pas une alternative viable. Le PowerPC 405 est quant à lui un processeur RISC 32 bits développé par IBM et mis en marché en 1998. Intégré à la série Virtex 2 Pro, son utilisation ne nécessite aucune ressource supplémentaire tant que les instructions peuvent être contenues dans 16 Ko de mémoire. Cadencé à l'horloge de l'engin de coïncidence, c'est-à-dire 50 MHz, ce processeur *System on Chip* peut alors exécuter 50 millions d'instructions par seconde. Intégré à une architecture en pipeline, ce processeur disposerait donc de 50 instructions par coïncidence triple. En considérant sa latence de synchronisation avec le VHDL et la lenteur de certaines de ses opérations, le PowerPC pourrait quand même permettre l'économie de quelques ressources en exécutant une minime partie du traitement total à effectuer. En revanche, après la série Virtex 4 Pro, le PowerPC n'est plus intégré au sein des FPGA. Puisque le développement pour la prochaine génération du scanner LabPET utilise des FPGA de la série Virtex 5 Pro, l'utilisation d'un PowerPC causerait un problème de compatibilité pour toute recherche future sur ce nouvel appareil. Par conséquent, le PowerPC 405 n'est pas considéré comme une option valide malgré son coût nul en ressources.

Conformément à ce qui précède, les ressources dédiées à un ou plusieurs des sous-processus de l'architecture en pipeline ne peuvent être affectées aux deux types de processeurs facilement disponibles dans le Virtex 2 Pro. L'ensemble de la récupération des coïncidences triples sera une solution composée uniquement de VHDL synthétisé. De plus, puisque dans un FPGA le respect des contraintes temporelles dépend principalement du nombre d'unités logiques totales utilisées, leur utilisation devra être la plus limitée possible. Effectivement, il est d'usage général de ne pas dépasser 90% d'utilisation totale pour ne pas avoir de problèmes de synchronisation des données. Par conséquent, l'intégration de la récupération des coïncidences triples ne devra pas dépasser ce seuil.

En conclusion, une fois intégré au scanner LabPET, le module effectuant la récupération en temps réel des coïncidences triples les traitera sans les envoyer à la collimation électronique. De manière à satisfaire les contraintes de performance et de ressources, ce dernier devra d'ailleurs être uniquement en VHDL synthétisé mais devra limiter le plus possible le nombre d'unités logiques utilisées.

CHAPITRE 6

Conception de la récupération des coïncidences triples

6.1 Architecture

L'architecture finale de la récupération des coïncidences triples, illustrée à la figure 6.1, présente un système en pipeline couplé à des blocs de contrôle. Le traitement d'une coïncidence triple est divisé en 7 étapes en série durant chacune au maximum 46 cycles d'horloge. À cet égard, le système respecte donc les contraintes de performance puisqu'il est capable de traiter un maximum de 1 087 000 triples par seconde à une horloge cadencée à 50 MHz.

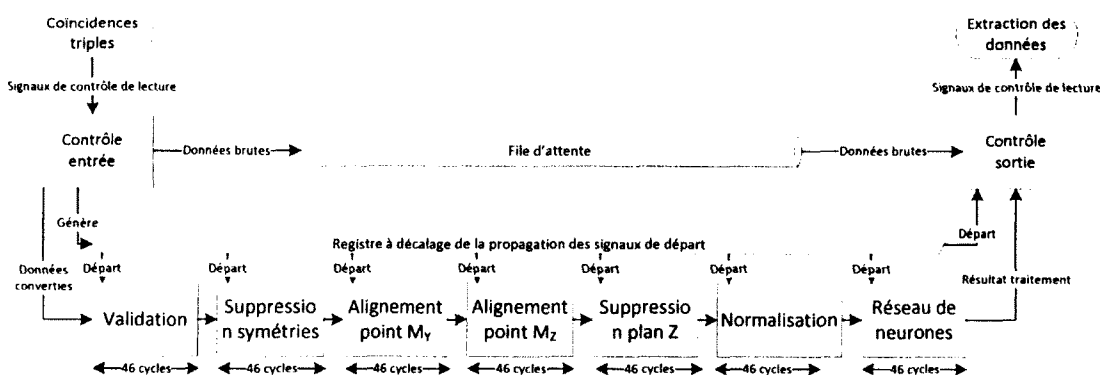


Figure 6.1 Architecture de la récupération des coïncidences triples

Après intégration dans l'engin de coïncidence du LabPET4, les contraintes de ressources sont elles aussi respectées tel que démontré dans le tableau 6.1.

Tableau 6.1 Ressources utilisés dans l'engin de coïncidence LabPET4

Ressource	Disp.	Avant	Ajout	Après	Utilisé
Unités logiques	23 616	12 340	5 463	17 803	75,4%
Blocs RAM 1 Ko	232	107	19	126	54,3%
Multiplicateurs 18 bit	232	0	45	45	19,4%

6.2 Représentation des données

Dans l'intention d'obtenir le traitement le plus performant possible à un coût minimal en ressources, l'ensemble des données transitant dans la récupération des coïncidences triples ont une représentation en **point fixe**. En effet, ce format est natif dans un FPGA et possède donc des avantages en terme de simplicité, de rapidité et d'économie de ressource comparativement à une représentation en point flottant. De plus, l'ensemble des calculs du système ne nécessite pas la précision accrue qu'aurait pu apporter le point flottant.

Dans le but de faciliter l'intégration au Virtex 2 Pro, le nombre de bits a été fixé à **18**. L'architecture de ce FPGA propose des bus de 18 bits pour tous les multiplicateurs et les blocs de mémoire. Il n'est donc pas avantageux de dépasser ce seuil, car ce genre de ressources serait automatiquement doublé. Ensuite, puisqu'un registre utilise une unité logique par 2 bits, l'économie réalisée par une représentation plus petite n'est pas assez substantielle pour contrebalancer la perte de précision.

L'évaluation de l'étendue des données et transformations du système permet ensuite de déterminer la séparation entre la partie entière et la partie fractionnaire. Dans l'optique de maximiser la précision du prétraitement tout en respectant l'étendue nécessaire au RNA, une distinction de représentation a été insérée entre ces deux étapes. L'énumération des étendues requises pour chacune des étapes du prétraitement dans le tableau 6.2 permet de déterminer qu'au minimum 3 bits d'entiers sont requis pour le prétraitement. En ajoutant un bit d'entier additionnel pour éviter les rares débordements causés par la normalisation des données, la représentation signée Q4.14 a été choisie pour le prétraitement.

Tableau 6.2 Évaluation des étendues des étapes préalables au RNA.

Considération	Étendue requise	Bits d'entiers requis ¹
Géométrie plan transaxial (dm)	$[-0,88; 0,88]$	1
Géométrie plan axial (dm)	$[0; 1,2]$	2
Énergie (MeV)	$[0; 1,00]$	2
Angles (rad)	$[-\pi; \pi]$	3
Cosinus et Sinus	$[-1; 1]$	2
Arc tangente	$[-\pi; \pi]$	3
Translations (dm)	$[-1,76; 0]$	2
Rotations (dm)	$[-1,88; 1,88]$	2
Normalisations	$[-\infty; \infty]^2$	-

¹Incluant le bit de signe dans une représentation signée.

²Division par 0 possible dans de rares cas. Étendue habituelle d'environ $[-1, 1]$.

La même évaluation est répétée dans le cas du RNA, où le tableau 6.3 résume les étendues requises du RNA. Pour cette étape, la représentation signée Q6.12 a été choisie, car elle permet de respecter l'étendue des poids et biais obtenus par l'entraînement des recherches antérieures [53]. Le bit additionnel d'entier a été ajouté pour permettre une flexibilité dans le cas où un autre projet de recherche connexe souhaiterait entraîner de nouveau le RNA.

Tableau 6.3 Évaluation des étendues du RNA.

Considération	Étendue requise	Bits d'entiers requis ¹
Tangente hyperbolique	$[-1, 1]$	2
Poids et biais du RNA	$[-10, 10]$	5

Finalement, l'ensemble des caractéristiques des représentations choisies est énuméré dans le tableau 6.4.

Tableau 6.4 Caractéristiques des représentations de données choisies.

Représentation	Étendue	Résolution
Q4.14	$[-8, 8[$	2^{-14}
Q6.12	$[-32, 32[$	2^{-12}

6.3 Contrôle des entrées

L'ajout d'un multiplexeur à la sortie de la formation des coïncidences permet d'envoyer toutes les coïncidences triples vers le système de récupération plutôt qu'à la collimation électronique. À cet effet, une logique d'entrée est nécessaire pour la lecture et la préparation des données. L'introduction du multiplexeur implique que le module de récupération des coïncidences triples et la collimation électronique se partagent la même source de coïncidences, c'est-à-dire une file d'attente. En considérant le temps de traitement inférieur de la collimation électronique, une coïncidence triple doit être lue et retirée de la file d'attente le plus rapidement possible afin de ne pas introduire un temps mort dans l'engin de coïncidences. C'est pourquoi une première lecture, indépendante du statut de la récupération des coïncidences triples, emmagasine toutes les nouvelles coïncidences triples dans une nouvelle file d'attente. Pour éviter tout débordement, la nouvelle file d'attente est de la même grosseur (512 x 64 bits) que la précédente et nécessite donc deux blocs mémoires.

¹Incluant le bit de signe dans une représentation signée.

Les lectures des événements inscrits dans la file d'attente précédente sont espacées d'un minimum de 46 cycles d'horloge. Une fois convertie dans le format requis, la coïncidence triple est prête à être traitée et un nouveau traitement doit être déclenché. Pour ce faire, le signal représenté par un registre à décalage est inversé. La détection de fronts sur certains des bits du registre génère alors des signaux de départ pour les modules du traitement en pipeline. La figure 6.2 présente un exemple où la détection de deux fronts déclenche des signaux départ des deux premiers modules.

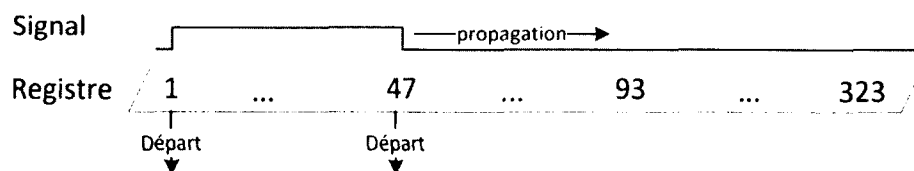


Figure 6.2 Registre à décalage de la propagation des signaux de départ.

Les coordonnées cartésiennes de la position des détecteurs sont inscrites au sein d'une LUT. En effet, la géométrie en cylindre du scanner permet de simplifier les données et de n'utiliser qu'un seul bloc mémoire plutôt que d'utiliser des fonctions trigonométriques et des multiplications. Le bloc mémoire (512 x 36 bits) contient donc les positions axiales (axe Z) des 16 anneaux et les positions transaxiales (plan XY) des 384 détecteurs de deux anneaux pour tenir compte du décalage angulaire. L'énergie en MeV est quant à elle obtenue par la multiplication par une constante de la valeur encodée (3 keV par bit). Cette étape ne nécessite donc qu'un seul multiplicateur. Parallèlement à la conversion des données, l'ensemble des paquets de données brutes représentant une coïncidence triple sont sauvegardés pour être réutilisés à la fin du traitement. Une nouvelle file d'attente (512 x 64 bits) consommant deux blocs mémoires est alors utilisée.

Tableau 6.5 Ressources utilisées par le contrôle des entrées.

Fonction	Unités logiques	Blocs RAM 1 Ko	Multiplicateurs
Lecture des données	142	4	0
Déclenchement de la récupération des coïncidences triples	198	0	0
Conversion des données	92	1	1
Persistance des données	68	2	0
Total	500	7	1

Le tableau 6.5 résume l'ensemble des ressources utilisées par le contrôle des entrées. À noter que les valeurs d'unités logiques présentées demeurent cependant approximatives, car à la synthèse du système complet certaines unités logiques peuvent être partagées. Cette affirmation est valable pour l'ensemble des modules présentés.

6.4 Validation d'une coïncidence triple

Cette étape, qui correspond au premier module en pipeline, détermine si une coïncidence triple respecte les conditions géométriques et énergétiques pour être incluse dans l'image. Dans l'état actuel du design, même si une coïncidence est jugée invalide, celle-ci est tout de même soumise au restant du traitement et sera écartée lors de la reconstruction de l'image.

$$(x_0 - x_1)^2 + (y_0 - y_1)^2 + (z_0 - z_1)^2 < 0,3025 \text{ dm}^2 \quad (6.1)$$

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 < 0,3025 \text{ dm}^2 \quad (6.2)$$

$$(x_2 - x_0)^2 + (y_2 - y_0)^2 + (z_2 - z_0)^2 < 0,3025 \text{ dm}^2 \quad (6.3)$$

Un triple est valide géométriquement uniquement lorsque deux événements sont rapprochés et que le dernier est isolé. En d'autres termes, un seul des trois vecteurs de distance doit être inférieur à 0,55 dm. En éliminant le calcul complexe d'une racine carrée, cette validation est équivalente aux équations 6.1 à 6.3. Outre la logique, trois multiplicateurs sont utilisés pour cette étape. À noter qu'après la validation géométrique, les événements au sein d'une coïncidence triple sont triés. L'événement isolé est toujours placé en premier, puis les deux événements regroupés sont triés de façon décroissante selon leur énergie. Ce tri ne nécessite seulement que des comparaisons de logique.

$$0,358 \text{ MeV} \leq e_p \leq 0,664 \text{ MeV} \quad (6.4)$$

$$0,358 \text{ MeV} \leq e_{s0} + e_{s1} \leq 0,664 \text{ MeV} \quad (6.5)$$

$$2 \times 0,358 \text{ MeV} \leq e_p + e_{s0} + e_{s1} \leq 2 \times 0,664 \text{ MeV} \quad (6.6)$$

La validation énergétique d'une coïncidence triple s'appuie sur les équations 6.4 à 6.6. L'énergie e_p du photon considéré photoélectrique, c'est-à-dire le photon isolé, doit avoir une énergie près de 0,511 MeV (équation 6.4). La somme des deux énergies e_{s0} et e_{s1} des deux photons considérés diffus, c'est-à-dire les deux photons regroupés, doit aussi respecter cette même condition (équation 6.5). Finalement, la somme totale en énergie doit être près de l'énergie totale dans une réaction d'annihilation (équation 6.6).

Tableau 6.6 Ressources utilisées par le module de validation.

Fonction	Unités logiques	Blocs RAM 1 Ko	Multiplicateurs
Validation géométrique	226	0	3
Tri énergétique	345	0	0
Validation énergétique	58	0	0
Total	629	0	3

Le module de validation a été soumis à 470 766 coïncidences triples provenant d’une simulation GATE. Le module en point fixe décrit dans cette section a pris la même décision en terme de validité qu’un autre programme équivalent en point flottant 470 748 fois (99.99%). Cette précision est jugée largement suffisante. Le tableau 6.6 résume d’ailleurs l’ensemble des ressources utilisées par le module de validation.

6.5 Simplification des données

Une fois validée, une coïncidence triple subit plusieurs transformations visant ultimement à éliminer sa composante Z. Parmi ces transformations, on retrouve entre autres quatre rotations par quaternion, chacune impliquant plusieurs opérations relativement complexes à effectuer dans un FPGA. Dans l’intention de respecter les contraintes de performance, chaque rotation est donc attribuée à un des quatre modules suivant la validation (voir la figure 6.1). Ces quatre modules, de la suppression des symétries jusqu’à la suppression du plan Z, sont très semblables et partagent les mêmes étapes principales.

6.5.1 Calcul de l’angle de rotation par arc-tangente

Le calcul des angles de rotation est effectué par la fonction trigonométrique arc-tangente. L’implémentation choisie pour cet opérateur est la technique itérative CORDIC [3] qui permet la rotation de vecteurs par des additions et décalages binaires. Cette méthode a été préférée à une LUT ou une approximation pour éviter une division et ainsi réaliser des économies en terme d’unités logiques. En revanche, cette technique est plus lente, car elle doit calculer itérativement les équations 6.7 à 6.9 pour atteindre une précision suffisante.

$$x_{i+1} = x_i - y_i \times d_i \times 2^{-i} \quad (6.7)$$

$$y_{i+1} = y_i + x_i \times d_i \times 2^{-i} \quad (6.8)$$

$$z_{i+1} = z_i - d_i \times \text{atan}(2^{-i}) \quad (6.9)$$

où $z_0 = 0$ et $d_i = +1$ lorsque $y_i < 0$, autrement -1

Pour les besoins de la rotation, la technique originale du CORDIC a été adaptée afin de couvrir la totalité du cercle trigonométrique. Cette adaptation sur 4 quadrants, communément appelée $atan2(y, x)$, permet d'obtenir l'angle en radians entre l'axe positif des abscisses et le point (x, y) (équation 6.10). Le calcul de l'angle requiert un peu plus de 300 unités logiques et dure 15 cycles d'horloge. Cette durée est d'ailleurs la cause de la séparation des rotations en quatre modules distincts.

$$atan2(y, x) = \begin{cases} atan(y, x) & x > 0 \\ atan(y, x) + \pi & y \geq 0, \quad x < 0 \\ atan(y, x) - \pi & y < 0, \quad x < 0 \\ \frac{\pi}{2} & y > 0, \quad x = 0 \\ -\frac{\pi}{2} & y < 0, \quad x = 0 \\ \text{Indéfini} & y = 0, \quad x = 0 \end{cases} \quad (6.10)$$

6.5.2 Rotation par quaternion

Puisque les quatre rotations sont toujours effectuées autour d'un seul axe référentiel (X, Y ou Z), les calculs reliés à des rotations par quaternions ou par matrice de rotation peuvent être grandement simplifiés. Après comparaison de ces deux techniques, on constate qu'elles nécessitent toutes deux des fonctions trigonométriques (cosinus et sinus) et qu'aucune n'a un avantage substantiel en termes de performance ou de coût en ressources. Cependant, la rotation par quaternions divise toujours par deux l'angle soumis aux fonctions trigonométriques. Pour cette raison, le choix s'est arrêté sur la technique des quaternions puisque l'étendue à couvrir des fonctions trigonométriques est réduite à $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Les ressources qu'aurait accaparées une étendue plus large peuvent alors être assignées de façon à obtenir une meilleure précision.

La simplification des quaternions, démontrée à l'annexe A, résulte en l'obtention des trois mêmes équations génériques (équations 6.11 à 6.13). Ces équations permettent alors de calculer la valeur résultante de la rotation par un angle α d'un point (x, y, z) autour d'un axe référentiel.

$$E1 = f \quad (6.11)$$

$$E2 = g - 2v^2g + 2wvh \quad (6.12)$$

$$E3 = h - 2v^2h - 2wvg \quad (6.13)$$

L'axe de rotation détermine alors l'assignation des variables f , g , h , w et v , puis l'assignation des équations aux composantes correspondantes du point final. Cette logique d'assignation, énumérée dans les tableaux 6.7 et 6.8, est conceptuellement représentée par de simples LUT.

Tableau 6.7 Assignation des entrées pour une rotation par quaternion

Axe	f	g	h	w	h
X	x	z	y	$\cos(\alpha/2)$	$\sin(\alpha/2)$
Y	y	x	z	$\cos(\alpha/2)$	$\sin(\alpha/2)$
Z	z	y	x	$\cos(\alpha/2)$	$\sin(\alpha/2)$

Tableau 6.8 Assignation des sorties pour une rotation par quaternion

Axe	x'	y'	z'
X	E1	E3	E2
Y	E2	E1	E3
Z	E3	E2	E1

Outre la logique nécessaire pour les assignations des entrées et des sorties, le calcul d'une rotation par quaternion se résume alors aux équations 6.12 et 6.13. La figure 6.3 présente l'architecture en pipeline conçue permettant de calculer ces équations à un coût de 165 unités logiques et 6 multiplicateurs.

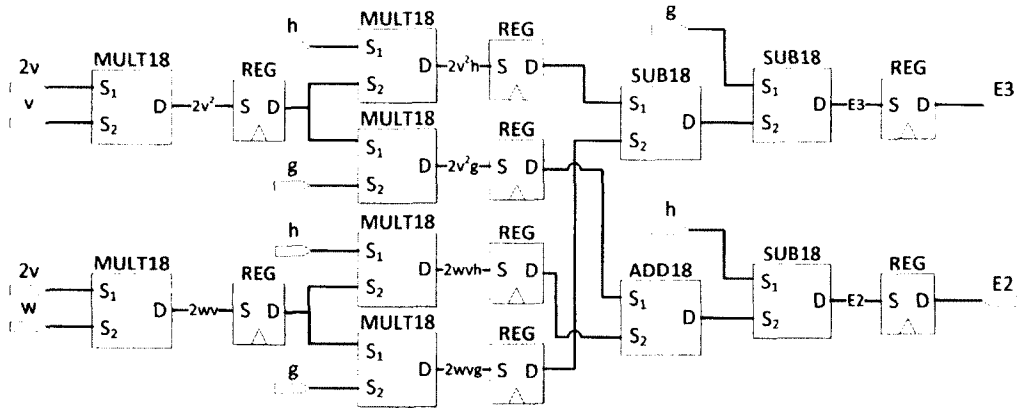


Figure 6.3 Calcul des équations 6.12 et 6.13 du quaternion.

6.5.3 Calcul des composantes de rotation par cosinus et sinus

Les composantes w et v des quaternions sont respectivement égales à $\cos(\alpha/2)$ et $\sin(\alpha/2)$ et doivent être évaluées avant le calcul des équations 6.12 et 6.13. En prenant en considération la lenteur relative du module d'arc tangente, ces fonctions trigonométriques doivent être le plus rapides possible. Dans ce contexte, l'utilisation d'une LUT est le choix désigné, car un CORDIC serait trop lent et les approximations linéaires sont plus coûteuses en ressources pour un rendement similaire en précision.

La représentation discrète de ces fonctions trigonométriques continues a été configurée dans le but de minimiser les intervalles entre les valeurs discrètes et ainsi optimiser la précision de la fonction. Tout d'abord, l'étendue a été réduite à $[-\pi/2, \pi/2]$ par la division de l'angle par deux des quaternions, puis diminué à $[0, \pi/2]$ en considérant l'équivalence de l'équation 6.14. Cette représentation uniquement positive permet d'ailleurs d'éviter à avoir à inclure un bit de signe. Ensuite, pour maximiser le nombre de valeurs discrètes possibles, le bloc de mémoire a été configuré à 2048 x 9 bits où chaque valeur a un bit d'entier et 8 bits fractionnels. Dans l'étendue $[0, \pi/2]$, cela ne nécessite cependant que 1608 valeurs. En conséquence, deux bits additionnels par valeur sont emmagasinés dans l'espace vacant et sont accessibles avec l'aide du deuxième port du bloc mémoire.

$$\cos(-\theta) = \cos(\theta) \text{ pour } \theta \text{ entre } [-\pi/2, \pi/2] \quad (6.14)$$

$$\cos(\theta) = \sin(\pi/2 - \theta) \quad (6.15)$$

L'équation 6.15 permet d'ailleurs d'utiliser le même bloc mémoire pour les fonctions cosinus et sinus. Cependant, puisque la précision additionnelle de deux bits accapare déjà le second port du bloc mémoire, les deux fonctions ne peuvent être calculées en même temps. Cette conception s'avère tout de même performante et économique, car le calcul d'une fonction ne requiert qu'un seul cycle d'horloge au coût de 37 unités logiques et un bloc mémoire.

6.5.4 Résultats

Chacun des modules effectuant une rotation a été testé individuellement et comparé aux résultats du programme en point flottant. Sur un échantillon de 100 000 points de rotations générés aléatoirement, les histogrammes obtenus des différences des résultats pour chacune des dimensions prennent la forme d'une courbe *cusp* dont les largeurs à mi-hauteur sont exprimées dans le tableau 6.9.

Tableau 6.9 Largeurs à mi-hauteur des histogrammes des différences entre les résultats en point flottant et en point fixe.

Module	X (mm)	Y (mm)	Z (mm)
Suppression des symétries	0,15	0,17	0
Alignement du milieu à $Y = 0$	0,07	0,04	0
Alignement du milieu à $Z = 0$	0,06	0	0,13
Suppression de la composante Z	0	0,02	0,02

Les histogrammes en forme de courbe *cusp* associés aux valeurs du tableau 6.9 sont quant à eux présentés dans les figures 6.4 à 6.11.

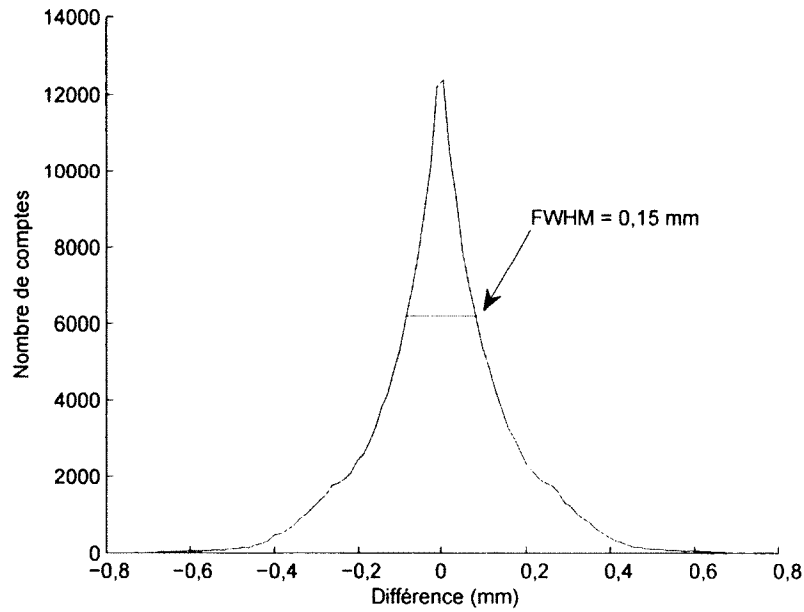


Figure 6.4 Suppression des symétries : histogramme des différences en X.

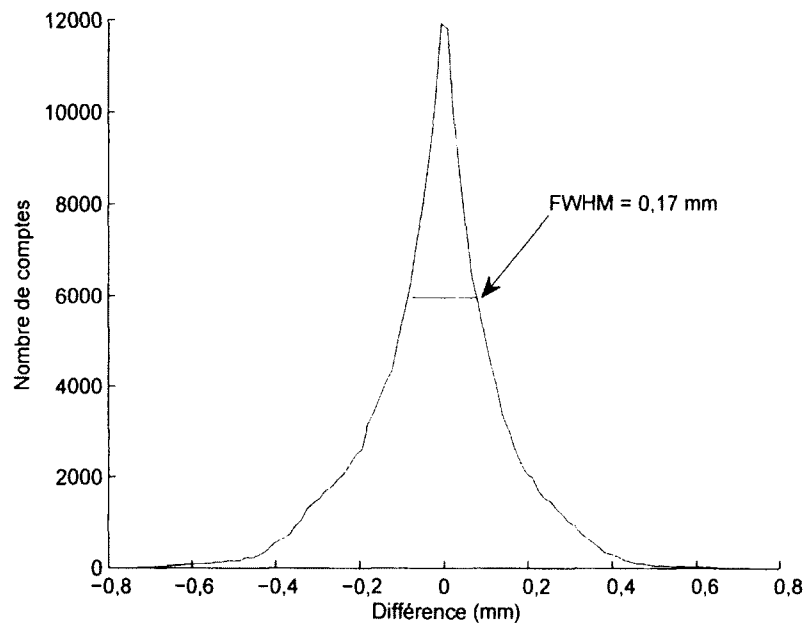


Figure 6.5 Suppression des symétries : histogramme des différences en Y.

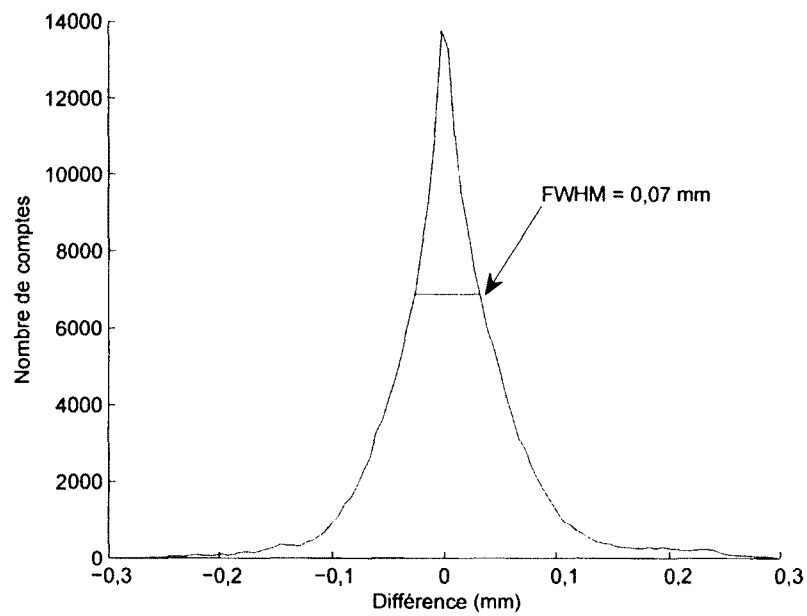


Figure 6.6 Alignement du point milieu à $Y = 0$: histogramme des différences en X.

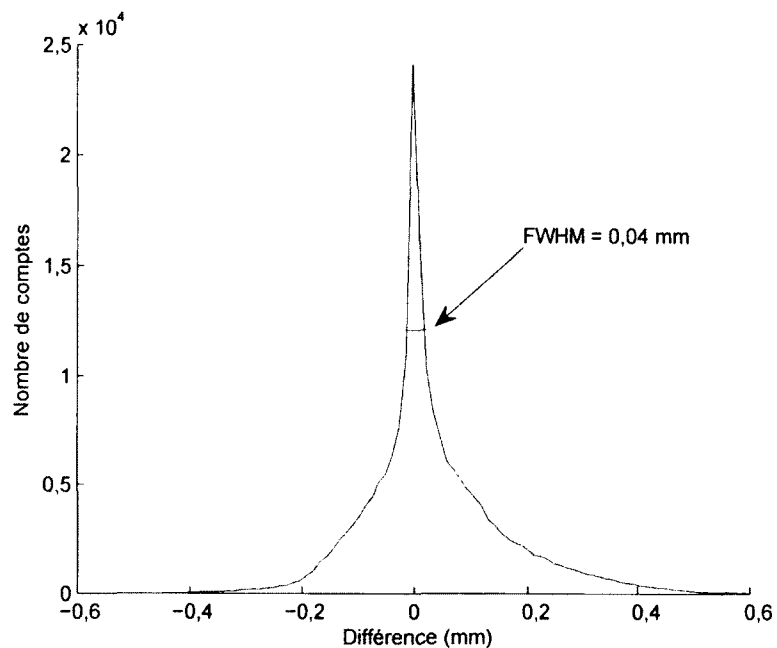


Figure 6.7 Alignement du point milieu à $Y = 0$: histogramme des différences en Y.

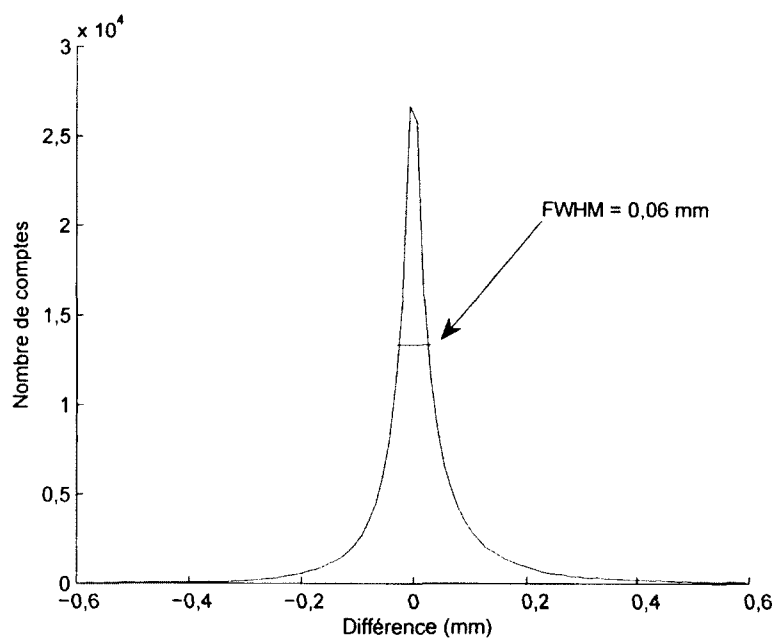


Figure 6.8 Alignement du point milieu à $Z = 0$: histogramme des différences en X.

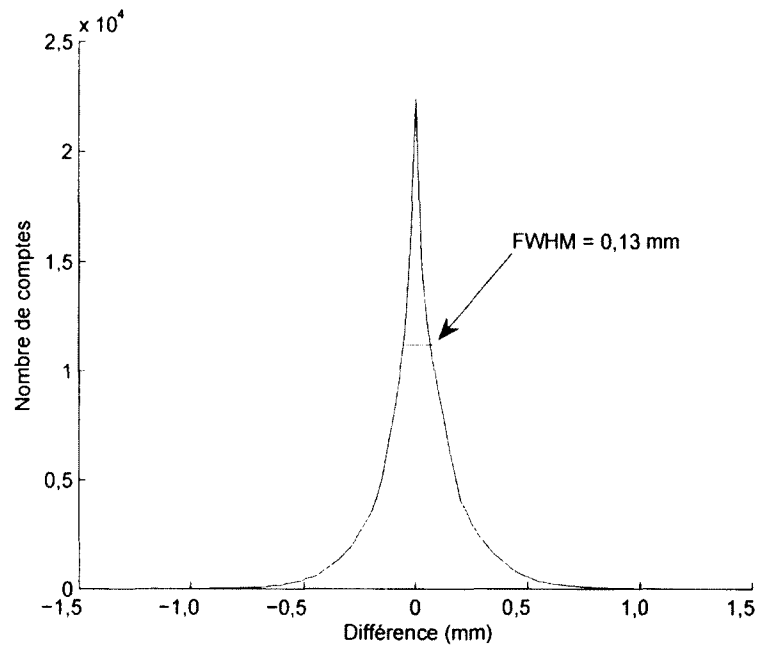


Figure 6.9 Alignement du point milieu à $Z = 0$: histogramme des différences en Z.

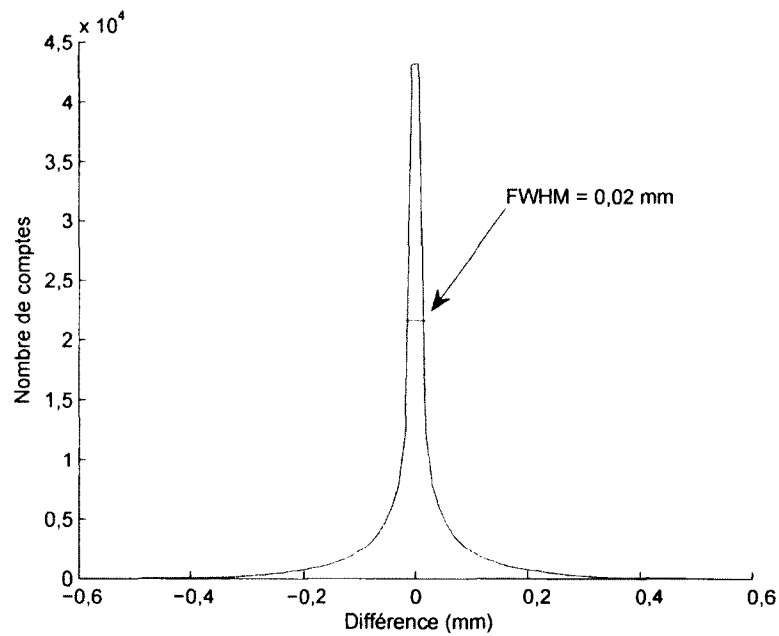


Figure 6.10 Suppression composante Z : histogramme des différences en Y.

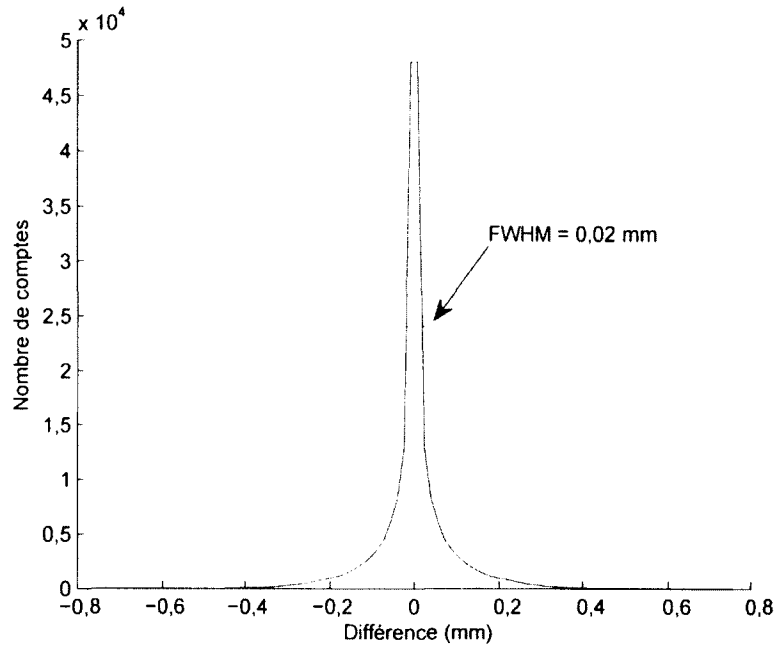


Figure 6.11 Suppression composante Z : histogramme des différences en Z.

Le tableau 6.10 énumère les ressources utilisées par chacune des fonctions principales dans la simplification des données.

Tableau 6.10 Ressources utilisées par les fonctions principales.

Fonction	Unités logiques	Blocs RAM 1 Ko	Multiplicateurs
Arc tangente	305	0	0
Cosinus et Sinus	57	1	0
Quaternion	165	0	6

Comme mentionné en début de section, chacune des rotations est attribuée à l'un des quatre modules servant à la simplification des données. Puisque ceux-ci partagent les mêmes étapes principales, leur coût en ressource est identique, hormis la suppression des symétries. Ce dernier module possède quelques étapes supplémentaires, comme une translation dans le plan Z. Représentées par des additions, ces étapes sont alors très simples à réaliser et ne valent pas la peine d'être décrites plus en détail. Les ressources utilisées par chacun des modules sont affichées dans le tableau 6.11.

Tableau 6.11 Ressources utilisées les modules de simplification des données.

Module	Unités logiques	Blocs RAM 1 Ko	Multiplicateurs
Suppression symétries	852	1	6
Alignement point M_y	810	1	6
Alignement point M_z	810	1	6
Suppression plan Z	810	1	6

6.6 Normalisation des données

La normalisation des données est la dernière étape avant le réseau de neurones artificiels. À cette étape, les deux photons considérés diffus ont subi plusieurs simplifications et leur composante Z a été éliminée. Le restant des informations, c'est-à-dire les coordonnées XY et l'énergie, doit être normalisé près de l'intervalle $[-1, 1]$ avant d'être présenté au RNA. Cette étape est essentielle, car l'entraînement du RNA s'est fait dans de telles conditions.

6.6.1 Multiplication inverse

Une division est requise pour chacune des 6 informations à normaliser. En considérant la haute disponibilité des multiplicateurs et la nécessité de minimiser le nombre d'unités logiques utilisées, la division est remplacée par la multiplication de l'inverse. Dans le cas de l'énergie, cette étape est simple puisque l'énergie est normalisée par rapport à une constante (511 keV). En revanche, la normalisation des coordonnées cartésiennes X et Y est plus complexe, car elle n'est pas effectuée par rapport à une constante, mais plutôt par rapport à une valeur dynamique dans l'intervalle $[-1, 1]$.

Afin de représenter la courbe du multiplicateur inverse qui a la forme d'une hyperbole décroissante, une approximation linéaire à intervalles a été choisie pour obtenir une meilleure précision que fournirait une simple LUT. Selon le diviseur en entrée, les paramètres a et b de l'intervalle concerné sont obtenus dans un bloc mémoire (512 x 36 bits) et permettent de calculer le multiplicateur inverse avec l'équation 6.16. En raison de la longueur du bloc mémoire et d'une logique permettant de gérer les diviseurs négatifs, les intervalles d'approximation sont longs de 2^{-9} et couvrent l'étendue $[0, 1]$.

$$m = a \times d + b \quad \text{où } m = \text{multiplicateur inverse et } d = \text{diviseur} \quad (6.16)$$

Néanmoins, les divisions près de zéro sont difficiles à représenter par leur pente importante. Dans les 65 premiers intervalles, plutôt que sauvegarder les paramètres d'approximation

a et b , le multiplicateur inverse m est directement sauvegardé en mémoire. La perte de précision pour ces intervalles est cependant jugée négligeable, car une analyse sur des données GATE révèle que ce genre de situation est rare et que son occurrence cause généralement un débordement dans la représentation Q6.12 du RNA. Par conséquent, une logique d'analyse de débordement permet de rejeter ces cas inutilisables.

6.6.2 Calcul de la normalisation

Une fois le multiplicateur inverse m obtenu, la normalisation des données peut être appliquée en effectuant le calcul des équations 6.17 à 6.19.

$$x_n = (x \times m) + 1 \quad (6.17)$$

$$y_n = (y \times m) \times 10 \quad (6.18)$$

$$e_n = z \times \frac{1}{0,511} \quad (6.19)$$

6.6.3 Résultats

Le module de normalisation a été testé individuellement et comparé aux résultats du programme en point flottant. Sur un échantillon de 65 000 données générées aléatoirement, les histogrammes obtenus des différences des résultats pour chacune des dimensions prennent la forme d'une courbe gaussienne dont les largeurs à mi-hauteur des coordonnées XY sont exprimées dans le tableau 6.12. Les graphiques de chacun des histogrammes sont quant à eux présentés dans les figures 6.12 et 6.13.

Tableau 6.12 Largeurs à mi-hauteur des histogrammes des différences entre les résultats en point flottant et en point fixe.

Module	X	Y
Normalisation	3×10^{-4}	7×10^{-4}

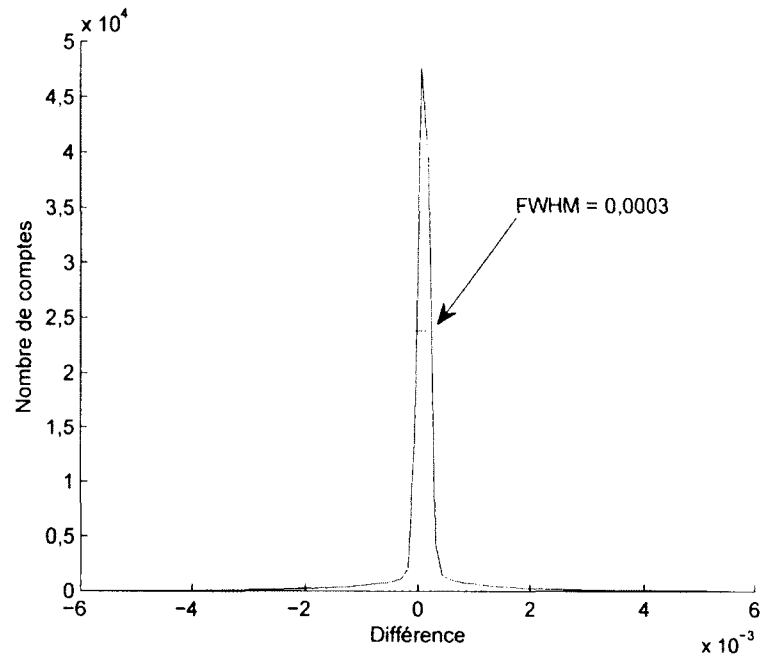


Figure 6.12 Normalisation : histogramme des différences en X.

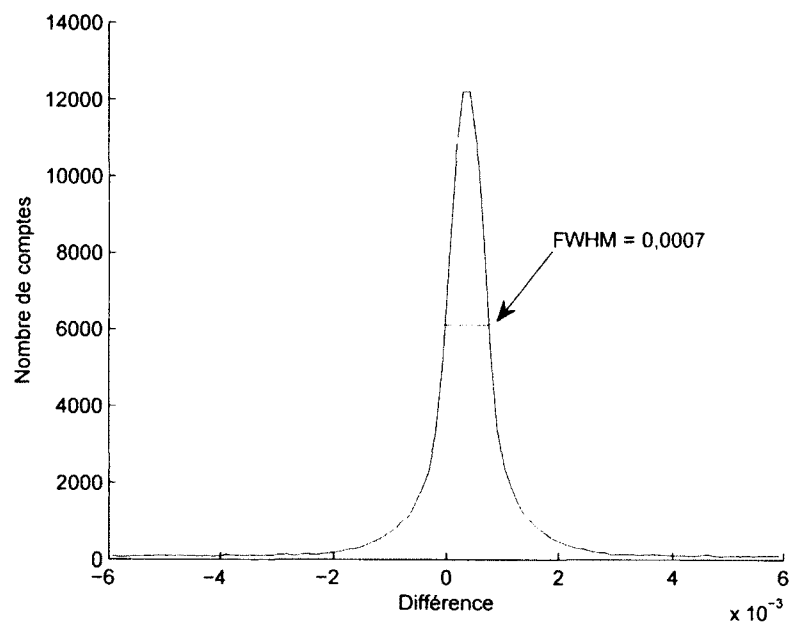


Figure 6.13 Normalisation : histogramme des différences en Y.

L'histogramme des différences pour la normalisation de l'énergie à la figure 6.14 résulte quant à lui en un bloc quasi-uniforme d'une largeur de 4×10^{-3} . Cette largeur est causée par la perte de précision pour représenter la constante (0,511 MeV). La représentation en point fixe de cette constante est d'ailleurs légèrement inférieure à celle du point flottant, d'où l'apparition de différences uniquement positives. Finalement, les oscillations sont quant à elles causées par la conversion sans arrondissement des résultats multiplicatifs en format Q6.12.

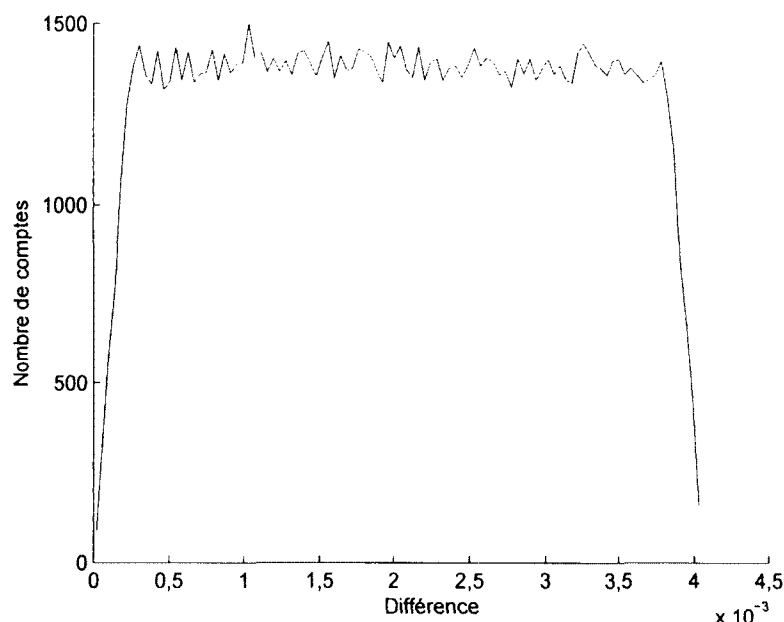


Figure 6.14 Normalisation : histogramme des différences en énergie.

Le tableau 6.13 résume l'ensemble des ressources utilisées par le module de normalisation.

Tableau 6.13 Ressources utilisées par la normalisation des données.

Fonction	Unités logiques	Blocs RAM 1 Ko	Multiplicateurs
Multiplicateur inverse	97	1	1
Calcul normalisation	377	0	8
Total	494	1	9

6.7 Réseau de neurones artificiels

Le réseau de neurones artificiels conçu possède autant de neurones que la plus large couche (10) et un engin de contrôle responsable de contrôler leurs entrées. Ce choix a été effectué

en fonction de la littérature qui indique que cette forme d'implémentation a un meilleur rapport performance versus ressource comparativement à ne concevoir qu'un seul ou tous les neurones. L'architecture obtenue est présentée à la figure 6.15. Ce schéma est simplifié visuellement en n'indiquant pas les ports d'horloge.

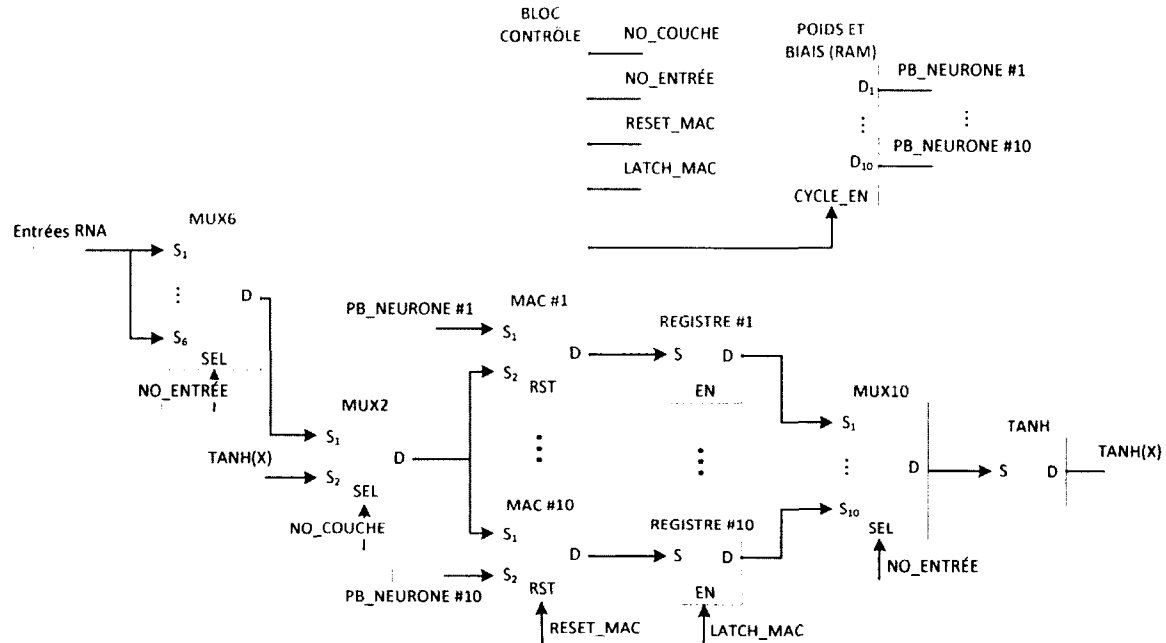


Figure 6.15 Architecture du réseau de neurones artificiels.

6.7.1 Blocs de calcul

Chaque neurone est conceptuellement représenté par une somme multipliée (MAC). À chaque cycle d'horloge, les deux entrées du MAC sont multipliées entre elles puis ajoutées à la somme. Un signal de réinitialisation permet de forcer la somme à zéro. Afin d'éviter tout débordement, la partie entière du registre contenant la somme a été agrandie. En considérant les limites de l'équation 6.20, l'obtention d'une grande valeur ne pose pas problème, car la somme est soumise à la fonction tangente hyperbolique.

$$\lim_{x \rightarrow -\infty} \tanh(x) = -1 \qquad \lim_{x \rightarrow +\infty} \tanh(x) = 1 \qquad (6.20)$$

Puisque l'assignation des entrées du MAC est contrôlée ailleurs, chaque neurone est alors dépourvu de logique. Par conséquent, chaque unité nécessite seulement un multiplicateur et 22 unités logiques pour le registre de somme.

6.7.2 Poids et biais

Les 115 poids et 16 biais du RNA, énumérés à l'annexe B, sont des constantes assignées à l'entrée des 10 blocs de calcul décrits dans la section précédente. Bien qu'un seul bloc mémoire est suffisant en espace, il ne permet pas d'accéder à 10 valeurs simultanément. Par conséquent, la même information est dupliquée dans trois blocs mémoire différents configurés 512 x 36 bits avec deux ports en sortie. Cette disposition permet de simuler 4 ports de sortie, car chaque espace mémoire de 36 bits contient deux poids ou biais de 18 bits. Finalement, puisque la séquence d'assignation des poids et biais est toujours la même, l'information dans les blocs mémoires est disposée selon cet ordonnancement. Une seule adresse de base décalée 9 fois permet par exemple d'obtenir 10 poids simultanés. Des signaux de contrôles extérieurs contrôlent alors l'incrémentement ou la réinitialisation de cette adresse de base.

6.7.3 Fonction d'activation tangente hyperbolique

Dans l'architecture présentée à la figure 6.15, bien que dix blocs de calcul soient présents, un seul module tangente hyperbolique est utilisé. Soit le résultat de la fonction tangente hyperbolique $t_{i,j}$ pour le neurone j de la couche i . Lorsque le résultat $t_{1,1}$ est soumis en entrée aux neurones de la deuxième couche, aucune autre valeur ne peut être soumise aux blocs de calcul MAC. Il n'est donc pas nécessaire d'avoir calculé simultanément les résultats $t_{1,2}$ à $t_{1,10}$. En conséquence, un seul module tangente hyperbolique est nécessaire, pourvu que celui-ci dispose de 10 registres pour conserver en mémoire les résultats des sommes multipliées. Dans cette architecture, le module tangente hyperbolique est contraint à ne durer qu'un seul cycle d'horloge. La fonction d'activation est alors représentée dans un bloc mémoire configuré 2048 x 9 bits agissant comme une LUT. La considération des équations 6.20 et 6.21 permet de ne couvrir uniquement que l'étendue $[0, 4[$ où l'intervalle entre les valeurs discrètes est de 2^{-9} . Une logique additionnelle permettra de résoudre les cas négatifs ou d'assigner à 1 toute valeur en entrée supérieure à l'étendue couverte. Le module tangente hyperbolique consomme finalement un bloc mémoire et 48 unités logiques pour sa simple logique et les 10 registres de sauvegarde.

$$\tanh(x) = -\tanh(-x) \quad (6.21)$$

6.7.4 Assignation des entrées des neurones

En considérant que les neurones d'une même couche reçoivent toujours les mêmes entrées, une logique précédant les blocs de calcul MAC sélectionne l'entrée courante destinée à être multipliée par un poids. Cette sélection est représentée par trois multiplexeurs différents (MUX2, MUX6 et MUX10) au sein de l'architecture du RNA à la figure 6.15. Le rôle de chacun des multiplexeurs dans la sélection des entrées est résumé dans le tableau 6.14.

Tableau 6.14 Multiplexeurs contrôlant les entrées des neurones.

Multiplexeur	Sélection
MUX2	Sélection de la source des entrées à appliquer aux neurones, c'est-à-dire choisir les entrées du RNA lors du calcul de la première couche ou la sortie du module tangente hyperbolique pour les couches subséquentes.
MUX6	Sélection de l'entrée courante parmi les entrées du RNA à soumettre aux neurones. La sortie du MUX6 est inutilisée suite au calcul de la première couche.
MUX10	Sélection du registre de somme multipliée à soumettre au module tangente hyperbolique. La sortie du MUX10 est inutilisée lors du calcul de la première couche.

6.7.5 Bloc de contrôle

Par sa méthode de conception, le RNA est un traitement séquentiel où les mêmes neurones seront utilisés à chacune des couches. La synchronisation du traitement est effectuée avec l'aide des signaux de contrôle générés par le bloc de contrôle. Celui-ci consiste tout simplement de compteurs incrémentés avec une logique générant les signaux de contrôle aux moments souhaités. Le tableau 6.15 résume d'ailleurs leur utilité au sein du RNA.

6.7.6 Résultats

Le réseau de neurones artificiels a été soumis à 470 748 coïncidences triples valides provenant d'une simulation GATE. Le RNA en point fixe décrit dans cette section a sélectionné le même photon diffus que le programme en point flottant 461 331 fois (97.99%). En considérant l'objectif du projet de recherche, qui vise principalement à déterminer si la technique de la récupération des coïncidences triples est faisable en temps, cette précision est jugée suffisante. L'optimisation du réseau, qui pourrait entre autres entraîner de nouveau le RNA dans des conditions en point fixe, est laissée à des itérations futures. Le

Tableau 6.15 Signaux de contrôle du RNA.

Signal	Définition	Utilisé par
NO_COUCHE	Représente la couche courante à calculer.	MUX2
NO_ENTRÉE	Représente l'indexe de l'entrée soumise aux neurones.	MUX6 et MUX10
RESET_MAC	Réinitialisation des sommes multipliées à la fin du calcul d'une couche.	Blocs de calcul MAC
LATCH_MAC	Mise en mémoire des sommes multipliées à la fin du calcul d'une couche.	Registres TANH
PB_CYCLE_EN	Itération des adresses de poids et biais du RNA pendant le calcul des sommes multipliées.	Blocs mémoires de poids et biais

résultat brut du réseau de neurones artificiels a aussi été comparé à celui du programme en point flottant. Sur un échantillon de 99 837 données générées aléatoirement, l'histogramme obtenu des différences des résultats prend la forme d'une courbe gaussienne dont la largeur à mi-hauteur est de $1,4 \times 10^{-3}$. Le graphique de cet histogramme est présenté à la figure 6.16.

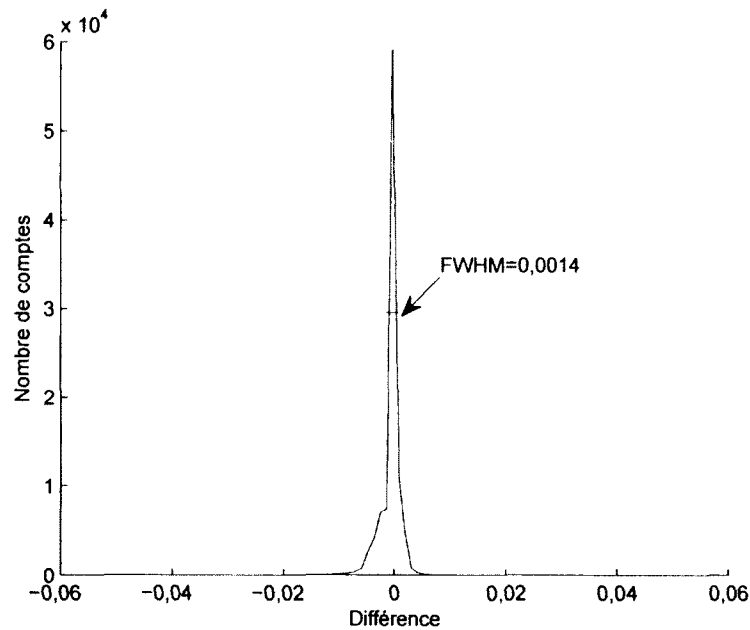


Figure 6.16 RNA : histogramme des différences.

Le tableau 6.16 résume l'ensemble des ressources utilisées par le réseau de neurones artificiels.

Tableau 6.16 Ressources utilisées par le réseau de neurones artificiels.

Fonction	Unités logiques	Blocs RAM 1 Ko	Multiplicateurs
Blocs de calcul MAC (10)	220	0	10
Mémoire poids et biais	27	3	0
Tangente hyperbolique ¹	138	1	0
Bloc de contrôle ²	363	0	0
Total	748	4	10

6.8 Contrôle des sorties

Conformément à l'insertion de la récupération des coïncidences triples tel que décrit à la section 5.6, la totalité des coïncidences triples traitées doit ensuite être transmise dans leur format original à l'extraction des données, et ce, quel que soit le résultat. Un bloc de contrôle des sorties est donc en charge de rétablir une coïncidence traitée dans son format original et d'y inscrire le résultat de la récupération. Les paquets de données brutes caractérisant une coïncidence triple transitant par le module de récupération sont toujours sauvegardés en mémoire en attendant la fin du traitement. Une logique démarrante après le RNA retire donc les paquets originaux de la mémoire interne et les sauvegarde ensuite dans une nouvelle file d'attente disponible pour l'extraction des données. Afin de permettre de reconstruire une image avec la récupération des coïncidences triples sans avoir à changer l'architecture de l'extraction des données, les résultats sont insérés dans l'octet du paquet servant habituellement au baseline (dont le format est décrit dans le tableau 5.1). Dans l'application actuelle du scanner, cette valeur est inutile et sa perte ne causera donc aucun problème. La nouvelle configuration du baseline est affichée au tableau 6.17 et la valeur de chacun des champs est expliquée plus en détail à l'annexe C.

Tableau 6.17 Reconfiguration du champ baseline.

7	7	6	6	5	5	4	2	1	0
Fortuit		Validité		Débordement		Inutilisé		Identification	

¹Incluant les registres de sauvegarde en entrée du module.

²Incluant les multiplexeurs responsables de l'assignation des entrées des neurones.

Le tableau 6.18 résume l'ensemble des ressources utilisées par le contrôle des sorties.

Tableau 6.18 Ressources utilisées par le contrôle des sorties.

Fonction	Unités logiques	Blocs RAM 1 Ko	Multiplicateurs
Rétablissement paquets	34	2	0
Insertion résultat	60	0	0
Total	94	2	0

En conclusion, grâce aux différents choix de conception présentés, le module de récupération en temps réel de coïncidences triples respecte les contraintes de performances et de ressources. Son effet sur l'image lors d'une acquisition reste cependant à être démontré.

CHAPITRE 7

Méthodologie

7.1 Acquisition des données

L'engin de coïncidences modifié qui inclut le module de récupération des coïncidences triples est testé sur le scanner LabPET4 pendant une durée 90 minutes. L'acquisition de données porte sur une mire ultra-micro-Derenzo [14] aux dimensions de 10 mm de long avec un rayon de 13,5 mm. Les différents points chauds (0,75, 1,0, 1,35, 1,7, 2,0 et 2,4 mm) de la mire sont remplis d'une solution aqueuse de ^{18}F FDG présentant une activité initiale de 8,63 MBq. Suite à cette acquisition, une normalisation d'une durée de 4 heures est effectuée sur une tige de germanium d'une activité de 9 MBq avec un diamètre de 6 cm. Le scanner est opéré en mode recherche avec filtration des événements à cristal inconnu. De manière à conserver le plus d'événements possible, aucune fenêtre en énergie et collimation électronique ne sont appliquées. Finalement, le tableau 7.1 montre les révisions utilisées pour chacune des composantes du scanner.

Tableau 7.1 Révisions utilisés pour l'acquisition.

Composante	Révision	Modifications apportées
Cartes d'acquisition	rev 4557	Aucune
DSP	rev 5315	Conversion énergie en keV
Cartes Hub	rev 4557	Aucune
Carte Com	rev 4557	Intégration de la technique de récupération des coïncidences triples
Control Panel	rev 5325	Aucune

7.2 Filtrage des données

Le fichier d'acquisition obtenu ne peut être directement utilisé pour une reconstruction d'image, car l'absence de fenêtre en énergie et de collimation électronique au moment de l'acquisition a causé l'apparition de coïncidences bruitées et inutilisables. De plus, les coïncidences triples sont toutes encapsulées dans un entête inconnu par le logiciel de reconstruction d'image. En conséquence, un logiciel programmé en *C#* permet d'analyser un tel fichier d'acquisition et d'en extraire l'information utile pour la reconstruction. Tout d'abord, afin de reproduire les conditions des recherches antérieures [52–55], un seuil minimal en énergie de 125 keV est appliqué sur tous les événements. Les coïncidences doubles restantes rencontrées subissent alors les validations suivantes :

1. Une validation du temps entre les deux événements limité à un maximum de 22 ns.
2. Une validation de l'énergie des événements en appliquant une fenêtre d'énergie entre 360 et 650 keV. Le seuil minimal peut aussi être configuré à 250 keV.
3. Une validation de la ligne de réponse au sein du champ de vue de 100 mm.
4. La création d'une coïncidence double si toutes les validations sont respectées.

Les coïncidences triples rencontrées sont quant à elles analysées selon plusieurs techniques. D'abord, la première analyse consiste de la récupération des triples en temps réel, où le résultat est déjà inscrit dans les paquets des événements. Ensuite, pour évaluer l'effet de l'absence de fenêtre d'énergie, les coïncidences au nombre d'événements supérieur à 3 qui n'ont pas été traitées en temps réel, mais qui après l'application du seuil de 125 keV fourniraient un triple sont évaluées par la même technique hors-ligne. Finalement, la technique de récupération sera comparée à une autre méthode déterminant que la bonne la ligne de réponse se retrouve entre les deux événements de plus haute énergie.

Pour simplifier la lecture du document, les coïncidences triples intégrées à l'image sont distinguées selon leur traitement. Les triples récupérés en temps réel sont libellés **TTR**, ceux hors-ligne **THL** et finalement les triples évalués selon leur énergie **TE**.

7.3 Reconstruction d'image

La reconstruction des images est effectuée avec l'aide des logiciels *TepData-Sino* et *Lab-PETrs64*. Le premier permet de créer les histogrammes nécessaires au second logiciel qui lui applique l'algorithme 3D MLEM sur le modèle du LabPET4. Une correction des coïncidences fortuites est appliquée, mais aucune correction de l'atténuation des coïncidences

diffuses n'est configurée. Le champ de vue est restreint à 46 mm comparativement au 100 mm utilisé lors du filtrage des données. Les images obtenues sont agrandies par interpolation trilineaire d'un facteur de 3 à l'aide du logiciel *Amide* [49]. Finalement, deux catégories d'image sont obtenues, c'est-à-dire des images de type isocompte et isolongueur.

7.3.1 Isocompte

Le terme isocompte réfère à des images où un nombre de comptes fixe permet de comparer l'effet des coïncidences utilisées sur l'image. Appliqué au contexte de recherche, les coïncidences doubles typiques (**D360**) sont comparées aux coïncidences doubles obtenues par abaissant du seuil minimal à 250 keV (**D250**) ainsi qu'aux coïncidences triples **TTR** et **TE**. De plus, une autre image composée de triples invalidés (**TI**) par le module de récupération des triples est générée pour confirmer que ces coïncidences rejetées sont bel et bien du bruit. Finalement, le nombre de coïncidences pour chacune de ces images est fixé à 20 millions, avant soustraction des coïncidences fortuites rencontrés à l'atteinte de cette valeur.

7.3.2 Isolongueur

Le terme isolongueur réfère à des images où, pour un temps fixe, est comparé le nombre de comptes obtenus. Appliquée au contexte de recherche, la totalité du fichier d'acquisition (90 minutes) est utilisée pour générer des images. Toutes les images décrites au tableau 7.2 incluent des coïncidences doubles et diffèrent par le type de coïncidences triples incluses. Deux ensembles d'images sont obtenus en variant le type de coïncidences doubles utilisées, soit les coïncidences doubles **D250** ou **D360**.

Image	Coïncidences incluses
1	Doubles
2	Doubles + TTR
3	Doubles + TTR + THL
4	Doubles + TE

Tableau 7.2 Coïncidences incluses dans chacune des images isolongueur.

7.4 Évaluation sur l'image

7.4.1 Sensibilité

L'évaluation de la sensibilité, uniquement effectuée sur les images isolongueur, vise à déterminer l'effet de la technique de la récupération des coïncidences triples sur le nombre de comptes total insérés dans l'image. Un rapport de sensibilité est donc attribué à chacune des images en comparant son nombre de comptes à celui de l'image du traitement original.

7.4.2 Contraste

L'évaluation du contraste vise à évaluer l'effet de la technique de la récupération des coïncidences triples sur l'arrière-plan de l'image. Le calcul du rapport CNR par l'équation 7.1 permet de calculer le rapport de contraste sur bruit entre deux régions de l'image : un point chaud de la mire (μ_h, σ_h) et une région rectangulaire d'arrière-plan (μ_b, σ_b). De plus, puisque la dimension axiale de ces régions ne couvre qu'une seule tranche d'image, tous les calculs de contraste doivent s'effectuer sur la même tranche transaxiale. Les deux régions sont d'ailleurs visuellement représentées en jaune sur la figure 7.1.

$$CNR = \frac{\frac{\mu_h - \mu_b}{\mu_b}}{\sqrt{\left(\frac{\sigma_h}{\mu_h}\right)^2 + \left(\frac{\sigma_b}{\mu_b}\right)^2}} \quad (7.1)$$

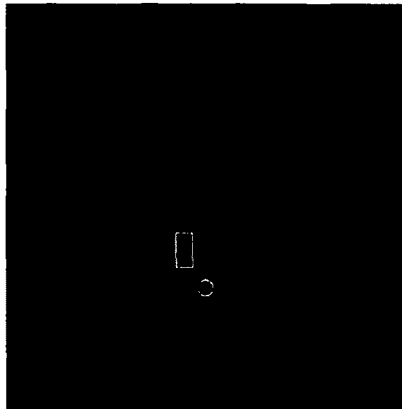


Figure 7.1 Régions utilisées pour l'évaluation du contraste.

Tel qu'illustré à la figure 7.2, le CNR est d'ailleurs évalué la cinquième itération de reconstruction, puis à chaque 10 itérations. Ce processus permet d'optimiser la qualité de l'image en sélectionnant l'image présentant le meilleur CNR. Toutes les images présentées au chapitre 8 ont été optimisées avec ce processus.

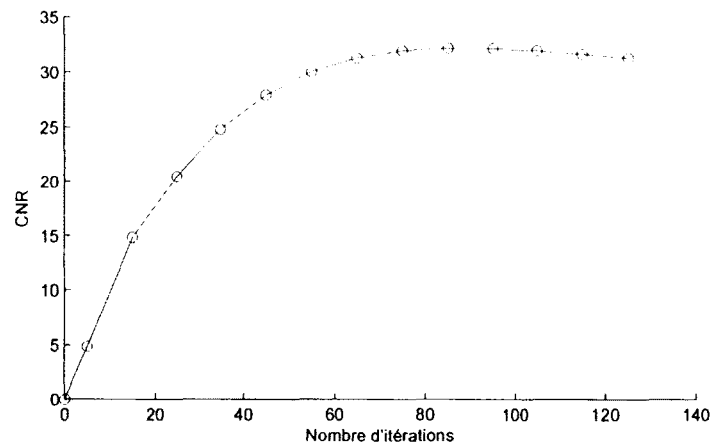


Figure 7.2 Sélection du CNR en fonction du nombre d'itérations.

7.4.3 Ligne de profil

L'évaluation de la ligne de profil est une autre façon de comparer la sensibilité et le contraste des images, mais aussi de comparer leur résolution. La ligne de profil, telle qu'illustrée à la figure 7.3, traverse deux points chauds de 2,4 mm puis trois autres de 1,7 mm. Toutes les lignes de profil sont calculées sur la même tranche transaxiale.

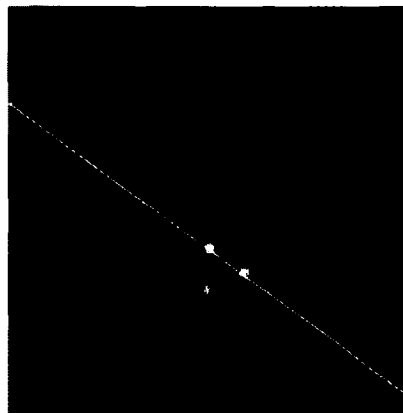


Figure 7.3 Ligne de profil utilisée.

En conclusion, le module de récupération des coïncidences triples est évalué grâce à une acquisition de données sur le scanner LabPET4. Les évaluations portent sur des images reconstruites à partir des données filtrées et ont comme critères d'évaluation la sensibilité, le contraste et la ligne de profil.

CHAPITRE 8

Résultats

8.1 Acquisition des données

L'analyse du fichier d'acquisition permet de déterminer la quantité de chacun des types de coïncidences avant l'application du filtrage des données. Ces valeurs, séparées selon la chaîne d'acquisition, sont énumérées au tableau 8.1.

Tableau 8.1 Total des coïncidences obtenues.

Coïncidence	Total prompts	Total fortuits
Double	594 037 549	35 712 836
Triple	69 086 312	10 492 921
Quadruple	6 241 974	2 038 639
Quintuple	472 647	301 916
Sextuple	49 919	36 595

8.2 Filtrage des données

L'application du filtrage des données permet d'extraire les coïncidences pouvant être utilisées à la reconstruction de l'image. De manière à comparer les résultats à ceux de [53], un seuil minimal d'énergie de 125 keV est tout d'abord appliqué sur tous les événements et le nombre de coïncidences restantes est affiché au tableau 8.2.

Tableau 8.2 Coïncidences après l'application d'un seuil de 125 keV.

Coïncidence	Total prompts	Total fortuits
Double	554 075 003	33 360 899
Triple	60 442 584	9 487 188

Le tableau 8.3 présente quant à lui le nombre de coïncidences pouvant être intégrées à l'image selon leur type après validation. Il faut cependant considérer que ces valeurs sont obtenues en validant la ligne de réponse dans un champ de vue de 100 mm. La reconstruction d'image, réalisée avec un champ de vue de 46 mm, est plus restrictive et diminuera bien évidemment le nombre de comptes final sur l'image.

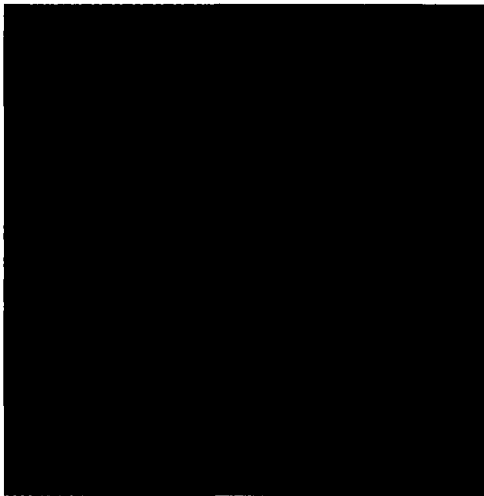
Tableau 8.3 Coïncidences pouvant être intégrées à l'image.

Type coïncidence	Prompts	Fortuits
Doubles 360 keV	59 442 438	2 966 285
Doubles 250 keV	114 051 430	6 702 664
Triples TTR	22 464 772	908 104
Triples THL	268 072	50 286
Triples TE	46 377 727	4 895 968

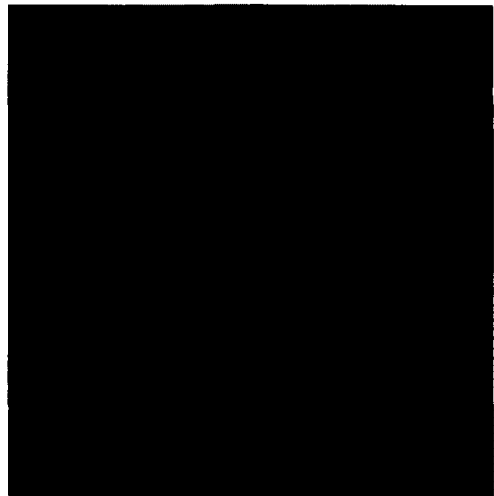
8.3 Reconstruction d'image

8.3.1 Isocompte

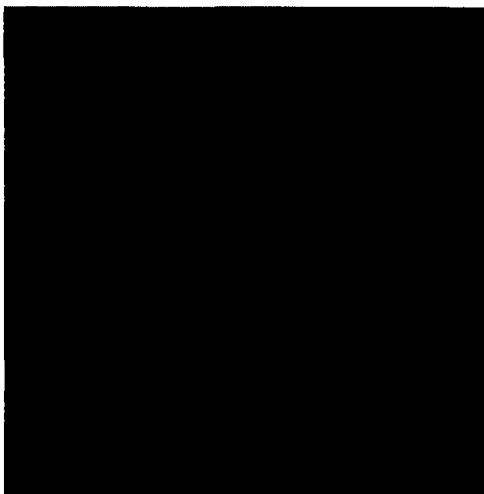
Les images suivantes correspondent aux images de type isocompte mentionnées à la section 7.3.1. Celles-ci sont reconstruites en utilisant 20 millions de coïncidences du type souhaité après soustraction des fortuits. Les valeurs plus détaillées du nombre de comptes et au contraste sont présentées à la section 8.4.



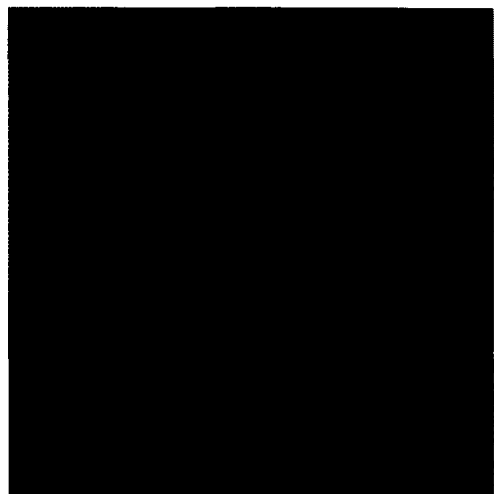
D250
19,3M coïncidences, CNR 30,3



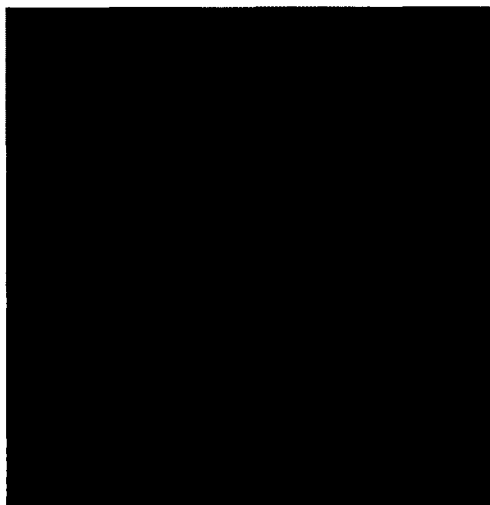
D360
19,4M coïncidences, CNR 20,0



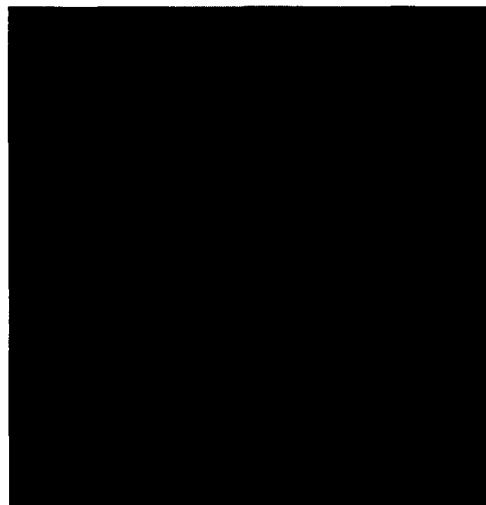
TTR
19,6M coïncidences, CNR 11,3



D360+TTR
19,4M coïncidences, CNR 25,4

**TI**

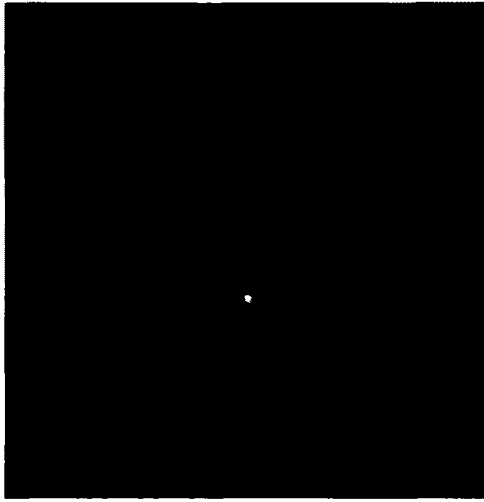
17,1M coïncidences, CNR 7,2

**TE**

18,4M coïncidences, CNR 10,2

8.3.2 Isolongueur

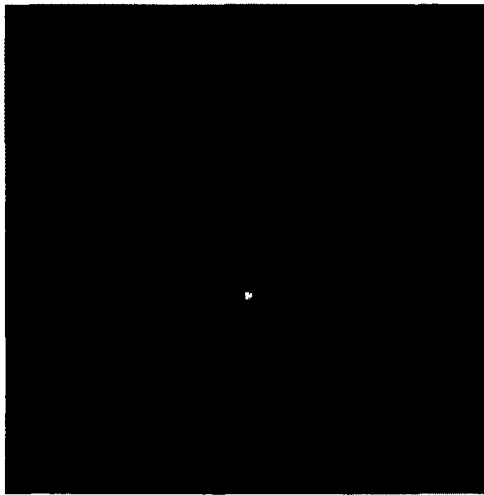
Les images suivantes correspondent à la reconstruction des deux ensembles de quatre types d'images isolongueur telles que décrites au tableau 7.2. Les valeurs plus détaillées du nombre de comptes et au contraste sont présentées à la section 8.4.



D250
101,0M coïncidences, CNR 21,2



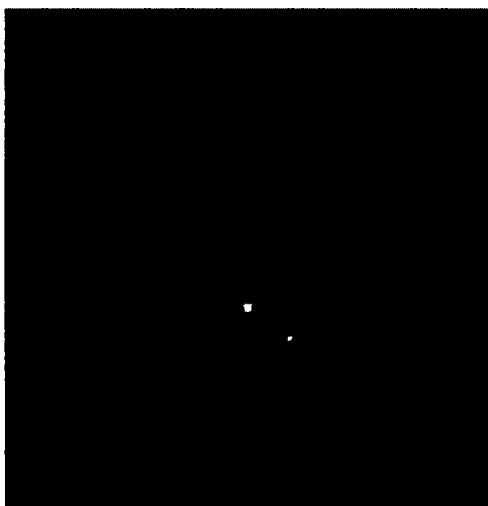
D360
51,7M coïncidences, CNR 32,2



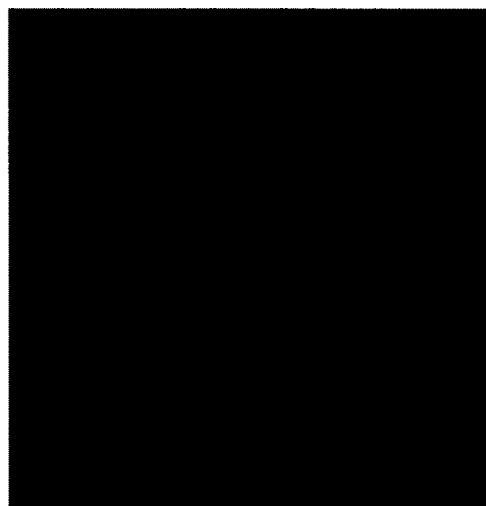
D250 + TTR
120,5M coïncidences, CNR 19,4



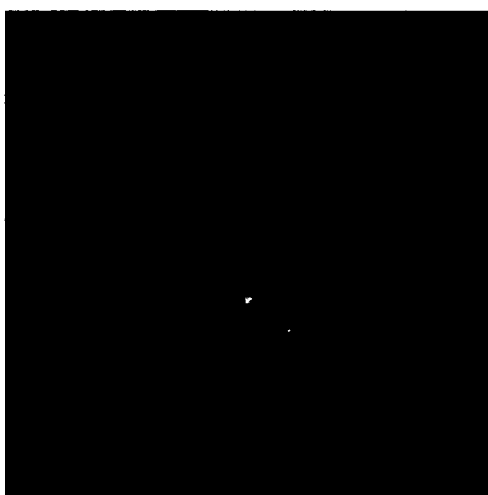
D360 + TTR
72,2M coïncidences, CNR 24,5



D250 + TTR + THL
120,6M coïncidences, CNR 19,3



D360 + TTR + THL
72,4M coïncidences, CNR 24,5



D250 + TE
138,8M coïncidences, CNR 18,0



D360 + TE
90,6M coïncidences, CNR 20,8

8.4 Évaluation sur l'image

8.4.1 Sensibilité

Pour chacune des images isolongueur, le tableau 8.4 énumère le nombre de comptes sur l'image ainsi que l'augmentation en pourcentage par rapport à l'image composée de doubles au seuil minimal de 360 keV. Ces valeurs incluent la soustraction des fortuits et la perte de certaines coïncidences avec l'application d'un champ de vue de 46 mm.

Tableau 8.4 Total des coïncidences obtenues.

Image	Total comptes	Gain (%)
D360	51 748 524	0
D360+TTR	72 200 446	39,7
D360+TTR+THL	72 409 325	39,9
D360+TE	87 438 805	69,0
D250	101 067 020	95,3
D250+TTR	120 485 746	132,8
D250+TTR+THL	120 694 625	133,2
D250+TE	135 705 568	162,2

L'inclusion des coïncidences triples **TTR** aux coïncidences doubles **D360** permet d'atteindre un gain en sensibilité de 39,7%. En présence d'une fenêtre d'énergie lors de l'acquisition, un gain de sensibilité additionnel de 0,2% aurait été possible. Comparativement à l'inclusion des triples **TE**, le gain en sensibilité est substantiellement inférieur. Néanmoins, il faut considérer que cette technique est beaucoup moins restrictive dans son inclusion des coïncidences, laissant présager une baisse de la qualité de l'image. Finalement, en connaissance des tableaux 8.2 et 8.3, le gain obtenu par l'abaissement de la fenêtre d'énergie à 250 keV est comme prévu bien supérieur au gain atteignable par l'inclusion des triples **TTR**.

8.4.2 Contraste

Le tableau 8.5 énumère les valeurs de contrastes des images de type isocompte ainsi que leur perte relative par rapport à l'image composée de coïncidences doubles **D360**. L'inclusion des coïncidences triples **TTR** résulte en la dégradation du CNR de 16,1%, valeur supérieure à celle obtenue dans l'image composée uniquement que de coïncidences doubles **D250**. Cependant, cette comparaison peut être nuancée en considérant la proportion des coïncidences triples **TTR** dans cette image. En effet, une image composée uniquement de genre d'événement produit une dégradation de 62,7% de la qualité de l'image, mais où il est toujours possible de distinguer les zones actives. Comme prévu, la qualité de l'image par rapport aux coïncidences triples **TE** est légèrement supérieure. Les valeurs de contrastes des images de type isolongueur affichées au tableau 8.6 permettent d'ailleurs de tirer les mêmes conclusions.

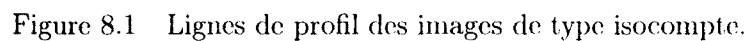
Tableau 8.5 Rapports de contraste sur bruit obtenus (Isocompte).

Image	CNR	Perte (%)	Nb itérations
D360	30,3	0	75
D250	20,0	34,0	75
TTR	11,3	62,7	65
D360+TTR	25,4	16,1	85
TI	7,2	76,2	65
TE	10,2	66,3	65

Tableau 8.6 Rapports de contraste sur bruit obtenus (Isolongueur).

Image	CNR	Perte (%)	Nb itérations
D360	32,2	0	85
D360+TTR	24,5	23,9	85
D360+TTR+THL	24,5	23,9	85
D360+TE	20,8	35,4	85
D250	21,2	34,1	85
D250+TTR	19,4	39,8	85
D250+TTR+THL	19,3	40,0	85
D250+TE	18,0	44,1	85

Les lignes de profil correspondant aux images de type isocompte sont affichées à la figure 8.1. Celles-ci confirment les résultats de contraste de la section 8.4.2, car le même ordonnancement est obtenu en triant les images selon leur concentration de comptes dans les points chauds.



Les lignes de profil correspondant aux images de type isolongueur sont affichées à la figure 8.2. Celles-ci confirment les résultats de sensibilité de la section 8.4.1, car le même ordonnancement est obtenu en triant les images selon la hauteur de leurs sommets dans les points chauds.

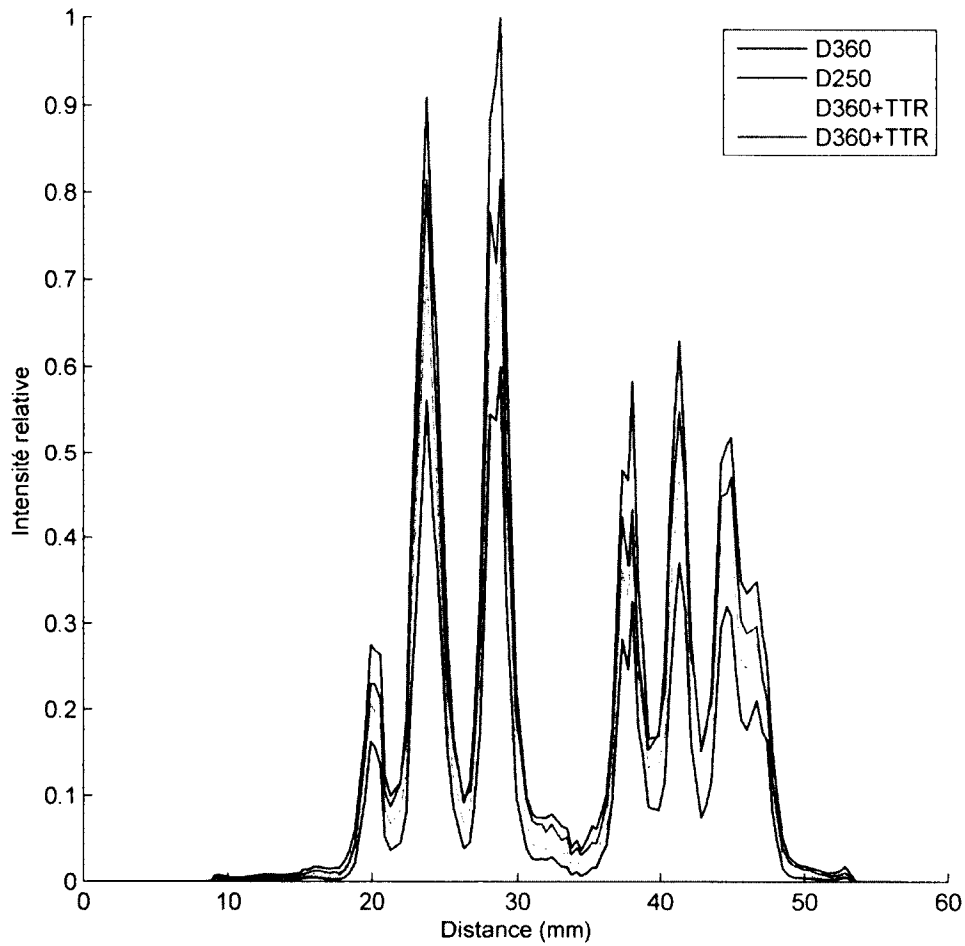


Figure 8.2 Lignes de profil des images de type isolongueur.

8.5 Discussion

8.5.1 Évaluation des résultats

Performance et ressources utilisées

Le système conçu respecte les contraintes de performances puisque celui-ci peut traiter 1 087 000 triples par seconde. De surcroît, la comparaison du module conçu à un programme en point flottant plus précis a révélé des résultats équivalents, où l'analyse de validité est la même dans 99,9% et où le choix de la ligne de réponse est la même dans 97,9% des cas. Par ailleurs, les objectifs liés aux contraintes de ressources sont eux aussi atteints, car le module de récupération des coïncidences triples a bel et bien été intégré à l'engin de coïncidence. En particulier, encore 25% des unités logiques demeurent disponibles à tout autre développement. En égard à ce qui précède, les ressources limitées d'un scanner de tomographie d'émission par positrons sont considérées suffisantes pour permettre l'implémentation d'un réseau de neurones artificiels devant traiter en temps réel les coïncidences triples du scanner.

En rétrospective, les contraintes de performances sont peut-être quelque peu excessives pour le LabPET4, car celles-ci sont basées sur une activité radioactive de 180 MBq (4.86 mCi). Cette dose est inconcevable dans une souris, en considérant que les acquisitions habituelles se situent autour de 37 MBq (1 mCi). Partant de ce fait et de la variation des fortuits en fonction de l'activité, cette activité pratiquement 5 fois plus élevée que la normale laisse croire que le nombre de triples estimé pourrait être 25 fois plus élevé. Dans l'éventualité où l'on ne souhaite pas imager de plus gros animaux (rat ou lapin), la performance accrue n'est probablement pas nécessaire et celle-ci pourrait être réduite en allouant plus de temps aux étapes en pipeline. De cette manière, certaines étapes pourraient être ralenties permettant ainsi l'économie de registres actuellement utilisés pour le respect des contraintes de temporelles. D'autre part, les étapes de rotations pourraient être regroupées et partager leurs ressources de façon à considérablement réduire leur coût en ressources.

Comparaison aux recherches antérieures

Les conclusions des recherches antérieures [53] ont tout d'abord déterminé que l'ajout de coïncidences triples à des coïncidences doubles (360 keV) permet d'améliorer la sensibilité de 54,2%. À priori, ce gain semble supérieur au gain de 39,7% obtenu dans ce projet de maîtrise. D'autant plus que les ratio de conversion de coïncidences triples brutes pouvant être récupérées (37,1% dans [53] versus 34,3% dans ce travail) sont très similaires.

Cependant, en comparant les résultats bruts de coïncidences triples, on remarque une différence entre les deux acquisitions sur le rapport coïncidences triples brutes et coïncidences doubles (15,8% dans [53] et 10,9% pour ce mémoire). Bien qu'une attention ait été portée pour acquérir les données du scanner dans des conditions semblables, il faut en conclure que certaines des conditions étaient différentes et ont amené cette variation. Afin de se convaincre de l'efficacité du RNA développé, une simple règle de 3 permet ainsi de déterminer qu'une acquisition fournissant de telles conditions aurait permis d'intégrer approximativement 9,3 millions de coïncidences triples supplémentaires à l'image avant l'application du champ de vue de 46 mm. Dans ces circonstances, il est valable de croire que la technique de récupération des triples en temps réel permet d'obtenir des résultats similaires à ceux obtenus précédemment au niveau de la sensibilité.

Les recherches antérieures [53] ont ensuite évalué qu'une reconstruction à partir de coïncidences triples permet d'obtenir une image où l'on distingue la plupart des points chauds, mais où la résolution est dégradée et un bruit d'arrière-plan est ajouté. L'ajout des coïncidences triples à des coïncidences doubles (360 keV) permet d'ailleurs d'obtenir un contraste similaire à celui obtenu si le seuil minimal d'énergie était réduit à 255 keV. Ces constats sont aussi valables dans le présent projet de recherche en comparant les valeurs de contraste.

En définitive, malgré sa précision inférieure due à sa représentation des données en point fixe, le module de récupération des coïncidences triples conçu dans le présent projet de recherche est quand même en mesure d'obtenir des résultats près de ceux des recherches antérieures [53] en terme de sensibilité et de contraste. Pour cette raison, l'on peut considérer que le réseau de neurones artificiels est capable de classer avec succès des coïncidences triples en temps réel. En liant cette conclusion à celle de la section précédente, il est alors possible de proposer l'implémentation décrite dans ce document pour répondre à la question de recherche :

Comment est-ce qu'un réseau de neurones artificiels peut être en mesure d'effectuer une classification en temps réel des coïncidences triples d'un scanner de tomographie d'émission par positrons, malgré les ressources limitées de ce dernier ?

8.5.2 Acquisitions à activité radioactive plus intense

Lors de l'acquisition des données dans le cadre du projet de recherche, le scanner a été opéré en mode recherche et sans fenêtre d'énergie. Dans cette configuration, l'extraction des données était pratiquement à pleine capacité malgré le degré d'activité radioactive relativement faible (8.63 MBq initialement) de la source utilisée. Effectivement, de manière

à permettre la reconstruction d'images du type avant-après, de nombreuses données habituellement écartées ont été conservées. Puisqu'il n'est pas possible d'améliorer la vitesse de l'extraction des données, le flux de données extraites doit être grandement réduit de manière à permettre des acquisitions à plus forte activité. Autrement, le module conçu dans le cadre du projet de recherche est inutilisable, car l'extraction des données sera engorgée.

La récupération des coïncidences triples doit être en conséquence une option disponible lors de la configuration d'une acquisition sur le scanner. Son activation permettrait d'obtenir un fichier d'acquisition où il n'est pas possible de distinguer les coïncidences triples récupérées des coïncidences doubles originelles. Dans cet ordre d'idée, il serait alors possible d'appliquer plusieurs opérations visant à réduire le flux de données, car celles-ci avaient été mises de côté lors de la phase d'évaluation du projet de recherche :

- activation de la collimation électronique ;
- compression des paquets des données grâce au mode image ;
- élimination des coïncidences triples invalides ;
- envoi d'une coïncidence triple récupérée en tant que double et
- application d'une fenêtre d'énergie au seuil minimal près de 125 keV.

Malgré toutes ces actions, le seuil minimal en énergie est néanmoins abaissé par rapport à sa valeur habituelle (typiquement 360 keV) pour permettre l'analyse des triples. Ceci provoque évidemment un flux plus important de données susceptibles d'engorger l'extraction des données. Il serait donc intéressant d'évaluer la possibilité d'ajouter une seconde fenêtre d'énergie plus restrictive après la formation des coïncidences pour tout groupe non triple. Celle-ci permettrait alors de restreindre les coïncidences doubles à faible énergie, sans empêcher une coïncidence triple d'être analysée.

8.5.3 Optimisations de performance et de ressource

Chacun des modules inclus dans l'architecture du module de récupération des coïncidences triples pourrait être optimisé individuellement et leur permettre d'être un peu plus performants et moins coûteux. Cependant, d'autres mesures permettraient d'obtenir des résultats plus substantiels.

Réduction des données à traiter

Dans la version actuelle du module de récupération, toute coïncidence triple jugée invalide par l'étape de validation est quand même soumise au restant de la chaîne de traitement, car ses informations doivent être récupérées pour la reconstruction d'images. Cependant, en considérant les propositions de la section 8.5.2, il n'est plus nécessaire de conserver ces cas jugés invalides et de surcroît leur allouer du temps de traitement. En conséquence, l'étape de validation, qui est environ 5 fois plus rapide que le temps maximal alloué, pourrait être ressortie du traitement en pipeline et filtrer tous les cas invalides. Ainsi, le module de récupération serait uniquement consacré qu'aux coïncidences triples jugées valides. Sans effectuer aucun autre changement, la performance globale serait alors automatiquement améliorée en passant de 1 087 000 triples par seconde à 1 087 000 triples *valides* par seconde. Le cahier des charges du système est alors modifié et une nouvelle analyse s'impose. La réduction de la performance requise permettrait d'appliquer une des mesures de la section 8.5.1 en allouant plus de temps aux étapes en pipeline pour ainsi réduire les ressources utilisées.

Reconfiguration du réseau de neurones

À la lumière des résultats de la comparaison du traitement en point fixe à celui en point flottant, le réseau de neurones artificiels n'a pas eu à être entraîné de nouveau pour bien fonctionner dans un environnement en point fixe. En revanche, en portant une attention particulière sur les entrées du RNA, les coordonnées Y des photons diffus après transformation sont toujours équivalentes, mais de signe différent. Puisque cette redondance n'est peut-être pas nécessaire au réseau de neurones pour sa compréhension du problème, il pourrait être avantageux de tenter de réduire le nombre d'entrées du RNA. Toute réduction valable permettrait peut-être de réduire la complexité du RNA et ainsi la charge de calculs à réaliser. À titre d'exemple, les deux ensembles de coordonnées XY pourraient être remplacés par le vecteur entre ces deux points. En utilisant la norme et l'angle, en plus de peut-être réduire la complexité du RNA, cette approche permettrait en plus de simplifier la normalisation des données et divisant par rapport à une constante et non une valeur dynamique.

8.5.4 Compatibilité aux autres versions du LabPET

Le module de récupération des coïncidences triples a été optimisé en fonction des caractéristiques du scanner LabPET4, mais qu'en est-il de sa compatibilité avec les autres versions existantes du scanner ainsi que les développements futurs? Du point de vue de

l'engin de coïncidences, les principales différences entre les versions correspondent à la fréquence d'horloge utilisée ainsi qu'au flux de données à traiter.

Puisqu'ils possèdent respectivement le double et le triple de détecteurs, les scanner LabPET8 et LabPET12 ont bien évidemment un flux de données supérieur à celui du LabPET4. La fréquence d'horloge de leur engin de coïncidences respectif a donc été cadencée plus rapidement à 78 MHz et 100 MHz. Il est difficile d'évaluer à l'avance si l'intégration du module de récupération au sein d'une de ces deux versions respecterait ces contraintes temporelles plus restrictives, car cela n'est déterminé qu'au moment de la synthèse finale du VHDL. Il est cependant prévisible qu'advenant quelques changements, la compatibilité vers le LabPET8 soit rapidement atteignable, car la plupart des composantes du module de récupération respectent déjà la contrainte temporelle plus stricte. En revanche, le cahier de charges du système doit être analysé à nouveau, car aucune estimation du flux maximal de coïncidences triples n'est connue dans cette configuration. Les mêmes constats peuvent être associés au LabPET12, mais sa compatibilité est plus difficilement atteignable en raison de la contrainte temporelle encore plus restrictive.

Au niveau du scanner en développement, le LabPETII, il est encore trop tôt pour évaluer la compatibilité puisque son engin de coïncidence n'est pas encore terminé. Néanmoins, puisque mis à part les blocs mémoires, aucun module spécifique au Virtex 2 n'est utilisé, l'ensemble du code pour le module de récupération des coïncidences triples peut être réutilisé. De plus, la représentation des données sur 18 bits n'aurait pas à être changée puisque le Virtex 5 propose des multiplicateurs 25 x 18 bits et des blocs mémoires où la sortie est un multiple de 18. Finalement, la nouvelle géométrie des détecteurs ne pose pas de problème puisque l'algorithme utilisé supporte tout type de géométrie cylindrique.

CHAPITRE 9

Conclusion

9.1 Sommaire

Le projet de recherche s'inscrit dans un contexte d'imagerie moléculaire, où la modalité d'imagerie d'intérêt est la tomographie d'émission par positrons (TEP) appliquée en recherche sur les petits animaux. Afin de permettre l'observation de détails infimes, les plus récents développements sur ce genre de scanner ont constamment amélioré leur résolution spatiale, sans toutefois obtenir les mêmes progrès en terme de sensibilité. Parmi les méthodes tentées afin de combler cette lacune, la récupération de coïncidences triples à l'aide d'un réseau de neurones artificiels semble être une technique viable puisqu'elle permet d'améliorer la sensibilité de 54,2% au coût d'une dégradation du contraste de l'image d'environ 20% [53]. Ces résultats demeurent cependant théoriques et la faisabilité d'une telle approche en temps réel n'est pas garantie, car le flux de coïncidences d'un scanner est très important et ses ressources de calculs sont limitées.

Dans ces circonstances, le projet de recherche présenté avait pour objectif principal de concevoir et d'optimiser sur un FPGA un réseau de neurones artificiels avec les ressources limitées du scanner LabPETTM, de façon à ce qu'il soit suffisamment performant pour être capable de classifier en temps réel le flux coïncidences triples provenant de ce dernier. Plus précisément, la vitesse de traitement du module conçu doit être d'un million de coïncidences triples par seconde tout en utilisant moins de 25% des unités logiques du FPGA XC2VP50. De plus, l'intégration à l'image des coïncidences triples récupérées en temps réel doit fournir des résultats semblables sur l'image à ceux des recherches antérieures.

En terme de résultats, la capacité de traitement obtenue est 1 087 000 coïncidences triples par seconde en utilisant 23,1% des ressources d'unités logiques. Comparativement à un programme équivalent en point flottant, l'analyse de validité prend la même décision dans 99,9% des cas et la même ligne de réponse est choisie dans 97,9% des cas. Intégrées à l'image, les coïncidences triples permettent une augmentation de sensibilité de 39,7%, valeur qui est en deçà de celle obtenue des recherches antérieures [53], mais cependant expliquée par une différence au niveau de l'acquisition. Au niveau de la qualité de l'image, la dégradation du contraste obtenu est similaire à celle observée antérieurement.

À la lumière de ces résultats, les ressources limitées d'un scanner de tomographie d'émission par positrons sont considérées suffisantes pour permettre l'implémentation d'un réseau de neurones artificiels devant classifier en temps réel les coïncidences triples du scanner. Suite à cette conclusion, il est alors possible de fournir une implémentation capable de répondre à la question de recherche : « *Comment est-ce qu'un réseau de neurones artificiels peut être en mesure d'effectuer une classification en temps réel des coïncidences triples d'un scanner de tomographie d'émission par positrons, malgré les ressources limitées de ce dernier ?* »

9.2 Contributions

La contribution originale principale de ce projet est bien évidemment de confirmer la faisabilité en temps réel de la technique de récupération des coïncidences triples par réseau de neurones artificiels, et ce, malgré les ressources limitées d'un scanner de tomographie d'émission par positrons. De surcroît, l'approche de réseau de neurones telle que développée par les recherches antérieures devient maintenant une alternative concrète pour améliorer la sensibilité.

Le projet de recherche présente d'ailleurs une architecture pouvant s'intégrer facilement à l'engin de coïncidence du scanner LabPET4. En effectuant quelques modifications pour que la récupération des coïncidences triples soit une option à activer lors d'une acquisition, son utilisation dans un contexte de recherche serait rapidement envisageable. En effet, selon le contexte de recherche, l'on peut favoriser la qualité d'image ou le gain en sensibilité. Dans le dernier cas, l'augmentation de sensibilité pourrait aussi être utilisée de façon à réduire le temps d'acquisition du scanner ou de réduire la dose radioactive nécessaire. Une image de qualité diagnostique serait donc obtenue plus rapidement ou avec moins de ressources.

Dans une autre perspective, la réussite du projet de recherche contribue à faire connaître la technique des réseaux de neurones dans domaine de la tomographie d'émission par positrons. En effet, cette approche est une approche pertinente à considérer en guise d'alternative aux solutions traditionnelles.

9.3 Travaux futurs

L'alternative que proposent les réseaux de neurones artificiels pourrait être étendue à d'autre type de coïncidences. Est-ce qu'un gain supplémentaire en sensibilité peut être obtenu en s'intéressant aux coïncidences multiples à plus de trois événements? Autrement, serait-il possible d'estimer le nombre de coïncidences fortuites? Sur le plan de la

conception, le module conçu peut-il être adapté afin d'être intégré à une nouvelle version en développement du scanner ? Par rapport à la précédente version, le scanner LabPETII améliore entre autres sa résolution spatiale en minimisant la taille de ses détecteurs. Avec un seuil d'énergie abaissé, la minimisation et la multiplication des détecteurs (37 000) laisse bien évidemment présager un flux de coïncidence important que le module actuel n'a pas la capacité de traiter. Par ailleurs, le changement de géométrie peut aussi exiger un ré-entraînement du réseau, ou même un changement complet de son architecture. Heureusement, l'architecture de l'engin de coïncidences est basée sur celle de la version précédente, d'où la possibilité réutiliser le même modèle d'insertion que celui présenté dans ce projet de recherche.

ANNEXE A

Simplification des quaternions

Soit le quaternion [84] q :

$$q = w + \vec{v} \text{ où } \vec{v} = xi + yj + zk \quad (\text{A.1})$$

Les paramètres w , x , y et z sont calculés en fonction de l'angle de rotation α :

$$w = \cos\left(\frac{\alpha}{2}\right) \quad (\text{A.2})$$

$$x = y = z = \frac{\sin(\frac{\alpha}{2})}{\|\vec{v}\|} \quad (\text{A.3})$$

Pour effectuer la rotation d'un point P , celui-ci doit d'abord être converti en un vecteur à 4 dimensions où la dimension ajoutée est nulle :

$$P = (a, b, c) \rightarrow \vec{P} = \begin{bmatrix} a \\ b \\ c \\ 0 \end{bmatrix} \quad (\text{A.4})$$

En calculant l'inverse du quaternion q précédent, il est possible d'effectuer la rotation et obtenir le nouveau point (ou vecteur) P :

$$\vec{P}' = q\vec{P}q^{-1} \quad (\text{A.5})$$

Cette opération peut aussi être représentée par la multiplication de deux matrices :

$$\vec{P}' = \begin{bmatrix} w^2 + x^2 - y^2 - z^2 & 2xy - 2wz & 2xz + 2wy & 0 \\ 2xy + 2wz & w^2 - x^2 + y^2 - z^2 & 2yz - 2wx & 0 \\ 2xz - 2wy & 2yz + 2wx & w^2 - x^2 - y^2 + z^2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \\ 0 \end{bmatrix} \quad (\text{A.6})$$

En utilisant uniquement des quaternions unitaires, c'est-à-dire :

$$\|q\| = \sqrt{w^2 + x^2 + y^2 + z^2} = 1 \quad (\text{A.7})$$

L'équation A.6 se simplifie alors à :

$$\vec{P}' = \begin{bmatrix} 1 - 2y^2 - 2z^2 & 2xy - 2wz & 2xz + 2wy & 0 \\ 2xy + 2wz & 1 - 2x^2 - 2z^2 & 2yz - 2wx & 0 \\ 2xz - 2wy & 2yz + 2wx & 1 - 2x^2 - 2y^2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \\ 0 \end{bmatrix} \quad (\text{A.8})$$

Puisque les rotations sont toujours définies par rapport à une axe référentiel, il est intéressant d'examiner la simplification des équations précédentes. En prenant pour exemple une rotation selon l'axe X, les composantes sont égales à :

$$\begin{aligned} i &= \pm 1 \\ j &= k = 0 \end{aligned}$$

Par déduction :

$$\begin{aligned} x &= \pm \sin\left(\frac{\alpha}{2}\right) \\ y &= z = 0 \end{aligned}$$

L'équation A.8 devient alors :

$$\vec{P}' = \begin{bmatrix} a' \\ b' \\ c' \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 - x^2 & -2wx & 0 \\ 0 & 2wx & 1 - 2x^2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \\ 0 \end{bmatrix} \quad (\text{A.9})$$

Après isolation :

$$a' = a \quad (\text{A.10})$$

$$b' = b - 2x^2b - 2wxc \quad (\text{A.11})$$

$$c' = c - 2x^2c + 2wxb \quad (\text{A.12})$$

En répétant les étapes équivalentes pour des rotations autour des axes Y et Z, le même genre d'équations sont obtenues. Trois équations sont alors posées :

$$E1 = f \quad (\text{A.13})$$

$$E2 = g - 2v^2g + 2wvh \quad (\text{A.14})$$

$$E3 = h - 2v^2h - 2wvg \quad (\text{A.15})$$

Finalement, la valeur attribuée aux variables f , g , h , w et v variera selon l'axe de rotation :

Axe	f	g	h	w	h
X	x	z	y	$\cos(\alpha/2)$	$\sin(\alpha/2)$
Y	y	x	z	$\cos(\alpha/2)$	$\sin(\alpha/2)$
Z	z	y	x	$\cos(\alpha/2)$	$\sin(\alpha/2)$

Axe	x'	y'	z'
X	E1	E3	E2
Y	E2	E1	E3
Z	E3	E2	E1

ANNEXE B

Valeurs de poids et biais

Le tableau B.1 liste l'ensemble des valeurs de poids (P) et biais (B) utilisées dans le réseau de neurones artificiel avant discrétisation pour chacun des neurones (N) de toutes les couches (C) .

Tableau B.1 Valeurs de poids et biais

C	N	P0	P1	P2	P3	P4	P5	P6	P7	P8	P9	B
0	0	1,340	-0,082	0,029	-6,942	0,010	-0,009	-	-	-	-	-2,330
0	1	-1,541	0,128	0,003	-9,698	-0,157	0,007	-	-	-	-	4,339
0	2	4,181	2,5325	0,058	-5,199	-2,591	-0,033	-	-	-	-	1,019
0	3	3,420	1,188	0,317	7,897	-1,279	-0,335	-	-	-	-	-6,383
0	4	-2,301	1,297	0,0417	9,969	-1,293	-0,0401	-	-	-	-	-2,181
0	5	1,069	1,439	-1,114	-0,851	-1,608	1,084	-	-	-	-	0,428
0	6	-4,120	0,173	-0,418	-4,713	-0,211	0,459	-	-	-	-	2,308
0	7	-4,104	-1,587	-0,324	-4,200	1,607	0,312	-	-	-	-	5,574
0	8	2,992	-1,335	-0,178	9,125	1,242	0,137	-	-	-	-	-4,394
0	9	1,009	0,043	-0,109	-7,974	0,058	0,129	-	-	-	-	6,473
1	0	-2,392	0,973	-2,821	3,550	-0,348	-1,376	-1,755	0,223	0,518	1,580	0,863
1	1	1,056	0,750	-0,515	-0,090	-0,187	-0,093	-0,107	0,234	0,117	-1,085	-1,111
1	2	0,011	-3,116	1,892	2,215	-1,237	0,088	0,133	0,927	-2,556	-0,371	0,135
1	3	-2,326	-8,295	0,119	0,905	12,241	-0,5613	-0,324	-0,762	5,833	1,304	0,974
1	4	-0,166	-0,734	-0,086	-0,689	-0,499	-0,263	1,461	3,461	0,038	-0,467	-1,604
2	0	1,638	-1,281	-4,490	-5,605	-5,415	-	-	-	-	-	1,108

ANNEXE C

Valeur du champ Baseline

En référence au tableau 6.17, le tableau suivant résume l'ensemble des valeurs inscrites au sein du champ baseline :

Champ	Valeur	Définition
Fortuit	0	Coïncidence de base.
	1	Coïncidence provenant de l'estimation des fortuits.
Validité	0	La coïncidence triple est invalide.
	1	La coïncidence triple est valide.
Débordement	0	La normalisation a été réussie.
	1	La normalisation a débordé.
Identification	00	L'événement appartient à une coïncidence invalide.
	01	L'événement est considéré comme le photon photoélectrique et fait parti du double originel.
	10	L'événement est considéré comme le photon diffus faisant parti du double originel.
	11	L'événement est considéré comme le photon diffus ne faisant pas parti du double originel.

Tableau C.1 Valeurs du champ Baseline

LISTE DES RÉFÉRENCES

- [1] Aliaga, R., Martinez, J., Gadca, R., Sebastia, A., Benlloch, J., Sanchez, F., Pavon, N. et Lerche, C. (2006). Corrected position estimation in PET detector modules with multi-anode PMTs using neural networks. *IEEE Transactions on Nuclear Science*, volume 53, numéro 3, p. 776–783.
- [2] Ametamey, S., Honer, M. et P.A., S. (2008). Molecular Imaging with PET. *Chemical Reviews*, volume 108, numéro 5, p. 1501–1516.
- [3] Andraka, R. (1998). A survey of CORDIC algorithms for FPGA based computers. *Proceedings of the 1998 ACM/SIGDA*, p. 191–200.
- [4] Bao, Q., Newport, D., Chen, M., Stout, D. et Chatziioannou, A. (2009). Performance evaluation of the invecon dedicated pet preclinical tomograph based on the nema nu-4 standards. *Journal of Nuclear Medicine*, volume 50, numéro 3, p. 401–408.
- [5] Bergeron, M., Cadorette, J., Beaudoin, J., Lepage, M., Robert, G., Selivanov, V., Tétrault, M., Viscogliosi, N., Norenberg, J., Fontaine, R. et Lecomte, R. (2009). Performance evaluation of the LabPET APD-based digital PET scanner. *IEEE Transactions on Nuclear Science*, volume 56, numéro 1, p. 10–16.
- [6] Blake, J., Maguire, L., McGinnity, T., Roche, B. et McDaid, L. (1998). The implementation of fuzzy systems, neural networks and fuzzy neural networks using FPGAs. *Information Sciences*, volume 112, numéro 1-4, p. 151–168.
- [7] Bloomfield, P., Myers, R., Hume, S., Spinks, T., Lammertsma, A. et Jones, T. (1997). Three-dimensional performance of a small-diameter positron emission tomograph. *Physics in Medicine and Biology*, volume 42, numéro 2, p. 389–400.
- [8] Budinger, T. (1998). PET Instrumentation : What Are the Limits? *Seminars in Nuclear Medicine*, volume 28, numéro 3, p. 247–267.
- [9] Chatziioannou, A. F. (2002). Molecular imaging of small animals with dedicated PET tomographs. *European Journal of Nuclear Medicine and Molecular Imaging*, volume 29, numéro 1, p. 98–114.
- [10] Cherry, S., Shao, Y., Silverman, R., Meadors, K., Siegel, S., Chatziioannou, A., Young, J., Jones, W., Moyers, J., Newport, D., Boutefnouchet, A., Farquhar, T., Andreaco, M., Paulus, M., Binkley, D., Nutt, R. et Phelps, M. (1997). MicroPET : a high resolution PET scanner for imaging small animals. *IEEE Transactions on Nuclear Science*, volume 44, numéro 3, p. 1161–1166.
- [11] Cherry, S., Sorenson, J. et Phelps, M. (2003). *Physics in Nuclear Medicine*, 3^e édition. Saunders, 523 p.
- [12] Chinn, G., Foudray, A. et Levin, C. (2006). Accurately positioning and incorporating tissue-scattered photons into PET image reconstruction. Dans *IEEE Nuclear Science Symposium Conference Record*. volume 3. p. 1746–1751.

- [13] Conti, M. (2009). State of the art and challenges of time-of-flight PET. *Physica Medica*, volume 25, numéro 1, p. 1–11.
- [14] Corporation, D. S. (2013). *Ultra-Micro Phantoms*. http://www.spect.com/pub/Ultra_Micro_Hot_Spot_and_Defrise_Phantoms.pdf (page consultée le 2013-10-28).
- [15] Cybenko, G. (1989). Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals, and Systems*, volume 2, numéro 4, p. 303–314.
- [16] Derenzo, S., Moses, W. et Huesman, R. (1993). Critical instrumentation issues for 2 mm resolution, high sensitivity brain PET. Dans *Quantification of Brain Function*. Elsevier Science Publishers, Amsterdam, p. 25–37.
- [17] Derenzo, S., Weber, M., Bourret-Courchesne, E. et Klintenberg, M. (2003). The quest for the ideal inorganic scintillator. *Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment*, volume 505, numéro 1, p. 111–117.
- [18] Di Domenico, G., Zavattini, G., Moretti, E., Piffanelli, A., Giganti, M., Motta, A., Sabba, N., Uccelli, L., Benini, E., Duatti, A., Bolzati, C., Boschi, A. et Del Guerra, A. (2003). YAP-(S) PET small animal scanner : quantitative results. *IEEE Transactions on Nuclear Science*, volume 50, numéro 5, p. 1351–1356.
- [19] Dorenbos, P. et de Haas, J. (1995). Non-proportionality in the scintillation response and the energy resolution obtainable with scintillation crystals. *IEEE Transactions on Nuclear Science*, volume 42, numéro 6, p. 2190–2202.
- [20] Drezet, A., Monnet, O., Mathy, F., Montemont, G. et Verger, L. (2007). CdZnTe detectors for small field of view positron emission tomographic imaging. *Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment*, volume 571, numéro 1-2, p. 465–470.
- [21] Evans, R. (1955). *The Atomic Nucleus*. International series in pure and applied physics, McGraw-Hill, 972 p.
- [22] Goerick, C., Noll, D. et Werner, M. (1996). Artificial neural networks in real-time car detection and tracking applications. *Pattern Recognition Letters*, volume 17, numéro 4, p. 335–343.
- [23] Hansen, L. et Salamon, P. (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 12, numéro 10, p. 993–1001.
- [24] Haykin, S. (2009). *Neural Networks and Learning Machines*, 3^e édition. Prentice Hall, 906 p.
- [25] Heinrichs, U. et Pietrzyk, U. (2003). Design optimization of the PMT-ClearPET prototypes based on simulation studies with GEANT3. *IEEE Transactions on Nuclear Science*, volume 50, numéro 5, p. 1428–1432.

- [26] Himavathi, S., Anitha, D. et Muthuramalingam, A. (2007). Feedforward neural network implementation in FPGA using layer multiplexing for effective resource utilization. *IEEE Transactions on Neural Networks*, volume 18, numéro 3, p. 880-888.
- [27] Hoffman, E., Guerrero, T., Germano, G., Digby, W. et Dahlbom, M. (1989). PET system calibrations and corrections for quantitative and spatially accurate images. *IEEE Transactions on Nuclear Science*, volume 36, numéro 1, p. 1108-1112.
- [28] Hoffman, E., Huang, S., Phelps, M. et Kuhl, D. (1981). Quantitation in Positron Emission Computed Tomography : 4. Effect of Accidental Coincidences. *Journal of Computer Assisted Tomography*, volume 5, numéro 3, p. 391-400.
- [29] Hume, S., Gumm, R. et Jones, T. (1998). Pharmacological constraints associated with positron emission tomographic scanning of small laboratory animals. *European Journal of Nuclear Medicine and Molecular Imaging*, volume 25, numéro 2, p. 173-176.
- [30] Hutchins, G., Miller, M., Soon, V. et Receveur, T. (2008). Small Animal PET Imaging. *Institute for Laboratory Animal Research Journal*, volume 49, numéro 1, p. 54-65.
- [31] IBM (2013). *PowerPC 405-S Embedded Processor Core User's Manual*. [https://www-01.ibm.com/chips/techlib/techlib.nsf/techdocs/8493418A173C1AC200257559007274AC/\\$file/ppc405Sum_1_2_pub.pdf](https://www-01.ibm.com/chips/techlib/techlib.nsf/techdocs/8493418A173C1AC200257559007274AC/$file/ppc405Sum_1_2_pub.pdf) (page consultée le 2013-10-28).
- [32] Iniewski, K. (2005). *Electronics For Radiation Detection*, 1^{re} édition. CRC Press, 378 p.
- [33] Knoess, C., Siegel, S., Smith, A., Newport, D., Richerzhagen, N., Winkeler, A., Jacobs, A., Goble, R. N., Graf, R., Wienhard, K. et Heiss, W.-D. (2003). Performance evaluation of the microPET R4 PET scanner for rodents. *European Journal of Nuclear Medicine and Molecular Imaging*, volume 30, numéro 5, p. 737-747.
- [34] Kohlmyer, S., Miyaoka, R. et Lewellen, T. (1999). Evaluation of low energy threshold settings for PVI PET systems. *IEEE Transactions on Nuclear Science*, volume 46, numéro 6, p. 2141-2145.
- [35] Korzhik, M., Fedorov, A., Annenkov, A., Borissevitch, A., Dossovitski, A., Missevitch, O. et Lecoq, P. (2007). Development of scintillation materials for PET scanners. *Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment*, volume 571, numéro 1-2, p. 122-125.
- [36] Larobina, M., Brunetti, A. et Salvatore, M. (2006). Small Animal PET : A Review of Commercially Available Imaging Systems. *Current Medical Imaging Reviews*, volume 2, numéro 2, p. 187-192.
- [37] Lecomte, R. (2004). Technology challenges in small animal PET imaging. *Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment*, volume 527, numéro 1, p. 157-165.

- [38] Lecomte, R. (2009). Novel detector technology for clinical PET. *European Journal of Nuclear Medicine and Molecular Imaging*, volume 36, numéro Suppl 1, p. S69–S85.
- [39] Lecomte, R., Cadorette, J., Richard, P., Rodrigue, S. et Rouleau, D. (1994). Design and engineering aspects of a high resolution positron tomograph for small animal imaging. *IEEE Transactions on Nuclear Science*, volume 41, numéro 4, p. 1446–1452.
- [40] Lecoq, P. (2010). Molecular Imaging Challenges With PET. *IEEE Transactions on Nuclear Science*, volume 57, numéro 3, p. 1485–1491.
- [41] Lee, D., Gaffar, A., Cheung, R., Mencer, O., Luk, W. et Constantinides, G. (2006). Accuracy-Guaranteed Bit-Width Optimization. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, volume 25, numéro 10, p. 1990–2000.
- [42] Lee, D.-U., Cheung, R., Luk, W. et Villasenor, J. (2008). Hardware Implementation Trade-Offs of Polynomial Approximations and Interpolations. *IEEE Transactions on Computers*, volume 57, numéro 5, p. 686–701.
- [43] Leroux, J., Tétrault, M., Rouleau, D., Pepin, C., Michaud, J., Fontaine, R. et Lecomte, R. (2009). Time discrimination techniques using artificial neural networks for positron emission tomography. *IEEE Transactions on Nuclear Science*, volume 56, numéro 3, p. 588–595.
- [44] Levin, C. (2005). Primer on molecular imaging technology. *European Journal of Nuclear Medicine and Molecular Imaging*, volume 32, numéro 14, p. 325–345.
- [45] Levin, C., Foudray, A. et Habte, F. (2006). Impact of High Energy Resolution Detectors on the Performance of a PET System Dedicated to Breast Cancer Imaging. *Physica Medica*, volume 21, numéro Suppl 1, p. 28–34.
- [46] Levin, C. S. et Hoffman, E. J. (1999). Calculation of positron range and its effect on the fundamental limit of positron emission tomography system spatial resolution. *Physics in Medicine and Biology*, volume 44, numéro 3, p. 781–799.
- [47] Lewellen, T. K. (2010). The Challenge of Detector Designs for PET. *American Journal of Roentgenology*, volume 195, numéro 2, p. 301–309.
- [48] Lo, S., Chan, H., Lin, J., Li, H., Freedman, M. et Mun, S. (1995). Artificial Convolution Neural Network for Medical Image Pattern Recognition. *Neural Networks*, volume 8, numéro 7-8, p. 1201–1214.
- [49] Locning, A.-M. (2013). *AMIDE : Amide's a Medical Imaging Data Examiner*. <http://amide.sourceforge.net/> (page consultée le 2013-10-28).
- [50] MacDonald, L. et Dahlbom, M. (1998). Parallax correction in PET using depth of interaction information. *IEEE Transactions on Nuclear Science*, volume 45, numéro 4, p. 2232–2237.
- [51] Melcher, C. (2000). Scintillation Crystals for PET. *Journal of Nuclear Medicine*, volume 41, numéro 6, p. 1051–1055.

- [52] Michaud, J., Brunet, C.-A., Lecomte, R. et Fontaine, R. (2010). Results from neural networks for recovery of PET triple coincidences. Dans *IEEE Nuclear Science Symposium Conference Record*. p. 3085-3087.
- [53] Michaud, J., Brunet, C.-A., Lecomte, R. et Fontaine, R. (2013). Efficacité de détection en TEP : approche par intelligence artificielle.
- [54] Michaud, J., Brunet, C.-A., Rafecas, M., Lecomte, R. et Fontaine, R. (2007). Sensitivity in PET : Neural networks as an alternative to Compton photons LOR analysis. Dans *IEEE Nuclear Science Symposium Conference Record*. volume 5. p. 3594-3600.
- [55] Michaud, J., Rechka, S., Brunet, C.-A., Lecomte, R. et Fontaine, R. (2009). Monte Carlo results from neural networks as an alternative to Compton photons LOR analysis. Dans *IEEE Nuclear Science Symposium Conference Record*. p. 3909-3916.
- [56] Missimer, J., Madi, Z., Honer, M., Keller, C., Schubiger, A. et Ametamey, S. (2004). Performance evaluation of the 16-module quad-HIDAC small animal PET camera. *Physics in Medicine and Biology*, volume 49, numéro 10, p. 2069-2081.
- [57] Musavi, M., Chan, K., Hummels, D. et Kalantri, K. (1994). On the Generalization Ability of Neural Network Classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 16, numéro 6, p. 659-663.
- [58] Naidu, S., Zafiriou, E. et McAvoy, T. (1990). Use of Neural Networks for Sensor Failure Detection in a Control System. *IEEE Control Systems Magazine*, volume 10, numéro 3, p. 49-55.
- [59] NVIDIA (2013). *GeForce 9800 GX2 Specifications*. <http://www.geforce.com/hardware/desktop-gpus/geforce-9800gx2/specifications> (page consultée le 2013-10-20).
- [60] Ollinger, J. et Fessler, J. (1997). Positron-emission tomography. *IEEE Signal Processing Magazine*, volume 14, numéro 1, p. 43-55.
- [61] OpenGATE (2013). *Simulations of Preclinical and Clinical Scans in Emission Tomography, Transmission Tomography and Radiation Therapy*. <http://www.opengatecollaboration.org/> (page consultée le 2013-10-20).
- [62] Ortigosa, E., Canas, A., Ros, E., Ortigosa, P., Mota, S. et Diaz, J. (2006). Hardware description of multi-layer perceptrons with different abstraction levels. *Microprocessors and Microsystems*, volume 30, numéro 7, p. 435-444.
- [63] Phelps, M., Mazziotta, J. et Schelbert, H. (1986). *Positron emission tomography and autoradiography : principles and applications for the brain and heart*. Raven Press, 690 p.
- [64] Pratz, G. et Levin, C. S. (2009). Bayesian reconstruction of photon interaction sequences for high-resolution PET detectors. *Physics in Medicine and Biology*, volume 54, numéro 17, p. 5073-5094.

- [65] Pysz, M., Gambhir, S. et Willmann, J. (2010). Molecular imaging : current status and emerging strategies. *Clinical Radiology*, volume 65, numéro 7, p. 500–516.
- [66] Rafecas, M., Böning, G., Pichler, B., Lorenz, E., Schwaiger, M. et Ziegler, S. (2003). Inter-crystal scatter in a dual layer, high resolution LSO-APD positron emission tomograph. *Physics in Medicine and Biology*, volume 48, numéro 7, p. 821–848.
- [67] Rafecas, M., Torres, I., Spanoudaki, V., Mcelroy, D. et Ziegler, S. (2007). Estimating accidental coincidences for pixelated PET detectors and singles list-mode acquisition. *Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment*, volume 571, numéro 1-2, p. 285–288.
- [68] Raichle, M. E. (1983). Positron Emission Tomography. *Annual Review of Neuroscience*, volume 6, p. 249–267.
- [69] Rumelhart, D., Hintont, G. et Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, volume 323, numéro 6088, p. 533–536.
- [70] Savich, A. W., Moussa, M. et Areibi, S. (2007). The impact of arithmetic representation on implementing MLP-BP on FPGAs : a study. *IEEE Transactions on Neural Networks*, volume 18, numéro 1, p. 240–252.
- [71] Sorsa, T., Koivo, H. et Koivisto, H. (1991). Neural networks in process fault diagnosis. *IEEE Transactions on Systems, Man and Cybernetics*, volume 21, numéro 4, p. 815–825.
- [72] Stickel, J. et Cherry, S. (2005). High-resolution PET detector design : modelling components of intrinsic spatial resolution. *Physics in Medicine and Biology*, volume 50, numéro 2, p. 179–195.
- [73] Strother, S., Casey, M. et Hoffman, E. (1990). Measuring PET scanner sensitivity : relating countrates to image signal-to-noise ratios using noise equivalents counts. *IEEE Transactions on Nuclear Science*, volume 37, numéro 2, p. 783–788.
- [74] Tai, Y.-C. et Laforest, R. (2005). Instrumentation Aspects of Animal PET. *Annual Review of Biomedical Engineering*, volume 7, p. 255–285.
- [75] Tétrault, M.-A., Oliver, J., Bergeron, M., Lecomte, R. et Fontaine, R. (2010). Real time coincidence detection engine for high count rate timestamp based pet. *IEEE Transactions on Nuclear Science*, volume 57, numéro 1, p. 117–124.
- [76] Turkington, T. G. (2001). Introduction to PET instrumentation. *Journal of Nuclear Medicine Technology*, volume 29, numéro 1, p. 4–11.
- [77] Valentine, J., Rooney, B. et Li, J. (1998). The light yield nonproportionality component of scintillator energy resolution. *IEEE Transactions on Nuclear Science*, volume 45, numéro 3, p. 512–517.
- [78] Vancijk, C. (2003). Inorganic scintillators in medical imaging detectors. *Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment*, volume 509, numéro 1-3, p. 17–25.

- [79] Vassiliadis, S., Zhang, M. et Delgado-Frias, J. G. (2000). Elementary function generators for neural-network emulators. *IEEE Transactions on Neural Networks*, volume 11, numéro 6, p. 1438–1449.
- [80] Vogl, T., Mangis, J., Rigler, A., Zink, W. et Alkon, D. (1988). Accelerating the Convergence of the Back-Propagation Method. *Biological Cybernetics*, volume 59, numéro 4, p. 257–263.
- [81] Walker, R., Purnell, G., Jones-Jackson, L., Thomas, K., Brito, J. et Ferris, E. (2004). Introduction to PET Imaging with Emphasis on Biomedical Research. *NeuroToxicology*, volume 25, numéro 4, p. 533–542.
- [82] Weber, S. et Bauer, A. (2004). Small animal PET : aspects of performance assessment. *European Journal of Nuclear Medicine and Molecular Imaging*, volume 31, numéro 11, p. 1545–1555.
- [83] White, S. (1989). Applications of Distributed Arithmetic to Digital Signal Processing : A Tutorial Review. *IEEE ASSP Magazine*, volume 6, numéro 3, p. 4–19.
- [84] WolframAlpha (2013). *Quaternion*. <http://mathworld.wolfram.com/Quaternion.html> (page consultée le 2013-10-28).
- [85] Xiaoguang, L., Medhat, M. et Shawki, A. (2006). Arithmetic formats for implementing artificial neural networks on FPGAs. *Canadian Journal of Electrical and Computer Engineering*, volume 31, numéro 1, p. 31–40.
- [86] Xilinx (2013). *Virtex-II Platform FPGAs : Complete Data Sheet*. http://www.xilinx.com/support/documentation/data_sheets/ds031.pdf (page consultée le 2013-10-28).
- [87] Xilinx (2013). *MicroBlaze Processor Reference Guide*. http://www.xilinx.com/support/documentation/sw_manuals/xilinx13_2/mb_ref_guide.pdf (page consultée le 2013-10-28).
- [88] Yang, F. et Paindavoine, M. (2003). Implementation of an RBF neural network on embedded systems : real-time face tracking and identity verification. *IEEE Transactions on Neural Networks*, volume 14, numéro 5, p. 1162–1175.
- [89] Yonggang, W., Junwei, D., Zhonghui, Z., Yang, Y., Lijun, Z. et Bruyndonckx, P. (2011). FPGA Based Electronics for PET Detector Modules With Neural Network Position Estimators. *IEEE Transactions on Nuclear Science*, volume 58, numéro 99, p. 34–42.
- [90] Ziemons, K., Barbier, R., Brandenburg, G., Bruyndonckx, P., Choi, Y., Christ, D., Costes, N., Declais, Y. et Devroede, O. (2005). The ClearPETTM project : development of a 2nd generation high-performance small animal PET scanner. *Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment*, volume 537, p. 307–311.

